Achievable Rates of Opportunistic Cognitive Radio Systems Using Reconfigurable Antennas with Imperfect Sensing and Channel Estimation

Hassan Yazdani, Azadeh Vosoughi, *Senior Member, IEEE* Xun Gong, *Senior Member, IEEE*
University of Central Florida
E-mail: h.yazdani@knights.ucf.edu, azadeh@ucf.edu, xun.gong@ucf.edu

***Abstract*—We consider an opportunistic cognitive radio (CR) system in which secondary transmitter ($\mathrm{SU_{tx}}$) is equipped with a reconfigurable antenna (RA). Utilizing the beam steering capability of the RA, we regard a design framework for integrated sector-based spectrum sensing and data communication. In this framework, $\mathrm{SU_{tx}}$ senses the spectrum and detects the beam corresponding to active primary user's (PU) location. $\mathrm{SU_{tx}}$ also sends training symbols (prior to data symbols), to enable channel estimation at secondary receiver ($\mathrm{SU_{rx}}$) and selection of the strongest beam between $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ for data transmission. We establish a lower bound on the achievable rates of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link, in the presence of spectrum sensing and channel estimation errors, and errors due to incorrect detection of the beam corresponding to PU's location and incorrect selection of the strongest beam for data transmission. We formulate a novel constrained optimization problem, aiming at maximizing the derived achievable rate lower bound subject to average transmit and interference power constraints. We optimize the durations of spatial spectrum sensing and channel training as well as data symbol transmission power. Our numerical results demonstrate that between optimizing spectrum sensing and channel training durations, the latter is more important for providing higher achievable rates.**

***Index Terms*—Achievable rates, beam detection, beam selection, channel estimation, imperfect spectrum sensing, opportunistic cognitive radio system, optimal and sub-optimal transmit power, reconfigurable antennas, training and data symbols.**

I. INTRODUCTION

A. Literature Review

Cognitive radio (CR) technology improves spectrum utilization and fills the spectral holes, via allowing an unlicensed or secondary user (SU) to access licensed bands in a such way that its imposed interference on license holder primary users (PUs) is restricted [1]. CR systems are mainly classified as underlay CR and opportunistic (or interweave) CR systems. In underlay CR systems, SUs use a licensed frequency band simultaneously with PUs, as long as the interference caused by SUs and imposed on PUs stays below a pre-determined threshold [1]–[3]. While underlay CR systems do not require spectrum sensing to detect PUs' activities, they demand coordination between PUs and SUs to obtain channel state information (CSI), that is not always feasible. In opportunistic CR systems, SUs use a licensed frequency band during a time interval, only if that frequency band is not used by PUs. While opportunistic CR systems do not require coordination between PUs and SUs to acquire CSI corresponding to SU-PU link (and hence the system implementation is easier), they necessitate spectrum sensing to monitor and detect PUs' activities.

This research was supported by NSF under grants ECCS-1443942 and CNS-2006683

The CR literature mainly assume that SU has access to full CSI of all links for its operation. However, in practice, SU has access only to partial CSI, due to several factors including channel estimation error, mobility of PU or SU, and limitation of feedback channel. Partial (imperfect) CSI has deteriorating effects on the fundamental performance limits of CRs and should not be overlooked. We note that the impact of partial CSI on the performance of underlay and opportunistic CR systems are different, due to inherent distinctions between these two CR systems. For underlay CR systems, several researchers have studied the impact of imperfect CSI on the ergodic capacity [4]–[9] and symbol error probability [10]. In particular, references [4]–[6] focus on investigating the impact of imperfect CSI of $\mathrm{SU_{tx}}$–PU receiver ($\mathrm{PU_{rx}}$) link on the optimal transmit power of $\mathrm{SU_{tx}}$ that maximizes the constrained capacity of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link, where $\mathrm{SU_{tx}}$ cannot always satisfy the interference power constraint (due to partial CSI) and has to reduce its transmit power. The authors in [7], [8] considered the impact of partial CSI for both $\mathrm{SU_{tx}}$–$\mathrm{PU_{rx}}$ and $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ links on the CR system capacity. Different from [4]–[8], [9] discussed the trade-off between channel estimation accuracy and channel estimation duration (time). The authors in [9] studied the optimal transmit power of $\mathrm{SU_{tx}}$ and optimal channel estimation duration, such that the capacity of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link is maximized, subject to a constraint on interference power imposed on $\mathrm{PU_{rx}}$.

In opportunistic CR systems, spectrum sensing is necessary for detecting PUs' activities and protecting the PUs against harmful interference. In general, any spectrum sensing (signal in noise detection) technique is prone to errors, that can be described as mis-detection or false alarm probability [11], [12]. On the other hand, imperfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link due to channel estimation error (even under perfect spectrum sensing) has negative influence on the link capacity. Imperfect spectrum sensing exacerbates the negative effect of imperfect CSI on the link capacity. Hence, for opportunistic CR systems, one needs to study the combined impacts of imperfect spectrum sensing and imperfect CSI on the system performance. Such study presents new challenges, compared with studies that focus on understanding only the effect of imperfect spectrum sensing, when CSI is perfect (or vice versa), on the link capacity. To the best of our knowledge, there are only a few works that have considered the aforementioned combined effects in their system performance analysis [13]–[16]. For example in [13] $\mathrm{SU_{tx}}$ estimates the received power from PU during sensing-estimation time and monitors PU's activity. If the spectrum is sensed idle, $\mathrm{SU_{tx}}$ with its imperfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link, sends data to $\mathrm{SU_{rx}}$ with a fixed power. The authors showed that the constrained capacity of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link can be significantly enhanced (subject to a constraint on the detection probability), via optimizing sensing-estimation time. The authors in [14] considered a delay-sensitive CR system with a different setup, where after spectrum sensing at $\mathrm{SU_{tx}}$, $\mathrm{SU_{tx}}$ transmits at fixed powers and rates, where these fixed values depend on the result of spectrum sensing (i.e., the transmit power and rate corresponding to spectrum being sensed idle are different from those corresponding to spectrum being sensed busy). The authors optimized these fixed powers and rates such that the defined effective capacity is maximized, subject to average transmit power and buffer length constraints. The authors in [15] considered a related problem to [14], where the two data transmit power levels are given and instead two training power levels as well as training period are optimized to maximize the achievable rate. The work in [16] considered different levels of CSI corresponding to $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ and $\mathrm{SU_{tx}}$–PU links, and studied optimal transmit power levels of $\mathrm{SU_{tx}}$, such that the capacity of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link is maximized, where the optimized power levels depend on the level of CSI.

In the above cited works SUs are equipped with single antenna. Multiple antennas and in particular transmit beamforming techniques have been utilized to ameliorate the performance degradation due to the interference imposed on PUs for underlay CR systems [17]–[19] and opportunistic CR systems [20] with perfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link available at $\mathrm{SU_{tx}}$. The authors in [21] considered an opportunistic CR system, where $\mathrm{SU_{tx}}$ has a single antenna and $\mathrm{SU_{rx}}$ has multiple antennas and applies maximum ratio combining (MRC) technique, and studied the combined effects of spectrum sensing error and imperfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link at $\mathrm{SU_{tx}}$ on the system bit error rate (BER) performance. Optimal spectrum sensing time, channel estimation time, and $\mathrm{SU_{tx}}$ transmit power are obtained, such that BER is minimized, subject to average transmit and peak interference power constraints. We note that the benefits of multi-antenna techniques come at the cost of requiring an expensive and power-hungry radio frequency (RF) chain per antenna, which consists of digital-to-analog converters, filters, mixers, and amplifiers.

Alternatively, reconfigurable antenna (RA) is a low-complexity and low-cost antenna technology that provides benefits similar to those of multiple-antenna techniques with a very low cost hardware, since a RA has only one RF chain [22]–[24]. RAs enable efficient exploitation of spatial diversity (via dynamically adjusting radiation pattern and beam steering/scanning capability) for reliable spectrum sensing and data transmission in CR systems. They are also capable of changing their parameters to dynamically adjust their polarization, carrier frequency and bandwidth [22], [23], [25]. Utilizing their beam steering capability and low cost hardware advantage, RAs can pave the path to the next generation of CR wireless communication systems for a wide range of applications, including personal communications, emergency-response, cyber-physical systems, tactical wireless communications, and 5G wireless systems [26]. For instance, RAs are employed in [27], [28] to establish directional wireless links, combat significant path-loss, and reduce the number of RF chains in mmWave massive MIMO systems. For both underlay and opportunistic CR systems, RAs are used to increase signal-to-noise ratio (SNR) for transmission and reception of directional signals [29], enhance spectrum sensing [29]–[31], and limit interference to and from PUs [32], [33]. Motivated by the advantages of RAs, in our study we assume that $\mathrm{SU_{tx}}$ is equipped with an RA that has beam steering capability.

B. Knowledge Gap, Research Questions, and Our Contributions

To the best of our knowledge, our work is the first to consider the combined effects of spectrum sensing error and imperfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link on the achievable rates of an opportunistic CR system with a RA at $\mathrm{SU_{tx}}$. In our opportunistic CR system, $\mathrm{SU_{tx}}$ relies on the beam steering capability of RA to detect the direction of PU's activity and also to select the strongest beam for data transmission to $\mathrm{SU_{rx}}$. We assume $\mathrm{SU_{tx}}$ sends training symbols to enable channel estimation at $\mathrm{SU_{rx}}$, and employs Gaussian input signaling for transmitting its data symbols to $\mathrm{SU_{rx}}$. Also, $\mathrm{SU_{rx}}$ shares its imperfect CSI of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link with $\mathrm{SU_{tx}}$ through an error-free low-rate feedback channel.

Assuming that there are average transmit power constraint (ATPC) and average interference constraint (AIC), we provide answers to the following research questions: How does spectrum sensing error affect accuracy of detecting the direction of PU's activity, estimating $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ channel, and selecting the strongest beam for data transmission? How do training symbol transmission and beam detection error (error in obtaining the true direction of PU's activity) affect interference imposed on PU? How do the combined effects of spectrum sensing error and channel estimation error, as well as beam detection error and beam selection error (error in finding the true strongest beam for data communication to $\mathrm{SU_{rx}}$) impact the achievable rates for reliable communication over $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link? How do the trade-offs between spatial spectrum sensing time, channel training time, data transmission time, training and data symbol transmission powers affect the achievable rates? How can we utilize these trade-offs to design transmit power control strategies, such that the achievable rates subject to ATPC and AIC are maximized? Our main contributions follow:

1) Given this system model, we establish a lower bound on the achievable rates of $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ link, in the presence of both spectrum sensing error and channel estimation error. We formulate a novel constrained optimization problem, aiming at maximizing the derived lower bound subject to AIC and ATPC.

2) Our problem formulation takes into consideration the combined effects of imperfect spectrum sensing and channel estimation as well as the errors due to (i) incorrect detection of the

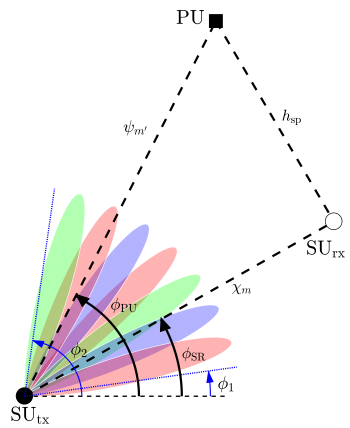


Fig. 1: Our opportunistic CR system with an M-beam RA at $\rm SU_{tx}$ and omni-directional antennas at $\rm SU_{rx}$ and PU.

beam corresponding to PU's location (and its corresponding effect on average interference imposed on PU) occurred during spatial spectrum sensing phase, (ii) incorrect selection of the strongest beam for data transmission from $\rm SU_{tx}$ to $\rm SU_{rx}$, occurred during channel training phase. These beam detection and beam selection errors are introduced by the RA at $\rm SU_{tx}$. 3) Given a fixed-length frame, we optimize the durations of spatial spectrum sensing and channel training as well as data symbol transmission power. Based on the structure of the optimized transmit power, we propose alternative power adaptation schemes that are simpler to implement and yield lower bounds on the achievable rates that are very close to the one produced by the optimized transmit power.

C. Paper Organization

The remainder of the paper is organized as follows. Section II explains our system model consisting of three phases: spatial spectrum sensing phase, channel training phase, and data transmission phase. Sections III and IV describe spatial spectrum sensing phase and channel training phase, respectively. Section V discusses data transmission phase, establishes a lower bound on the achievable rates, and characterizes ATPC and AIC. Then, it formalizes a constrained optimization problem with three optimization variables (durations of spatial spectrum sensing and channel training phases, and data symbol transmission power), aiming at maximizing the derived lower bound, subject to ATPC and AIC. Section VI provides solution to this constrained optimization problem. Section VII presents our simulation results and Section VIII concludes the paper.

II. System Model

A. Structure of a RA

We consider a RA which can generate M *beampatterns* and these beampatterns cover the angular plane from ϕ_1 to ϕ_2, i.e., the angular space from ϕ_1 to ϕ_2 is divided into M spatial *sectors* or beams[1]. One can extend this angular space to cover the entire azimuth plane. The beampattern corresponding to m-th beam achieves its maximum at angle $\kappa_m = \frac{2\pi(m-1)}{M}$ for $m = 1, \dots, M$. Fig. 1 shows the beampatterns of a RA with $M = 7$ beams. It is noteworthy that the RA can also reconfigure itself to generate an omni-directional pattern. To mathematically model the radiation pattern of beams, we adopt the Gaussian pattern in $x-y$ azimuth plane in terms of angle ϕ given by [32]

$$p(\phi) = A_1 + A_0\, e^{-B\left(\frac{\mathcal{M}(\phi)}{\phi_{3\rm dB}}\right)^2}, \qquad \mathcal{M}(\phi) = \text{mod}_{2\pi}(\phi+\pi)-\pi, \tag{1}$$

where $\text{mod}_{2\pi}(\phi)$ denotes the remainder of $\frac{\phi}{2\pi}$, $B = \ln(2)$, $\phi_{3\rm dB}$ is the 3-dB beamwidth, A_1 and A_0 are two constant antenna parameters. The radiation pattern of m-th beam at angle ϕ is

$$p_m(\phi) = p(\phi - \kappa_m), \qquad \text{for } m = 1, \dots, M. \tag{2}$$

In this paper, we discuss the received or transmitted signal at m-th beam of $\rm SU_{tx}$. This implies that, during the signal reception or transmission, the $\rm SU_{tx}$'s antenna parameters are set and tuned such that the beampattern corresponding to m-th beam is generated. Given the antenna design, we focus on how the sector-based structure of this RA can be exploited to enhance the system performance of our opportunistic CR system, in which $\rm SU_{tx}$ optimizes its sector-based data communication to $\rm SU_{rx}$ according to the results of its sector-based spectrum sensing.

B. Description of Our Opportunistic CR System

Our opportunistic CR system model is illustrated in Fig. 1, consisting of a PU and a pair of $\rm SU_{tx}$ and $\rm SU_{rx}$. We note that PU in our system model can be a primary transmitter or receiver. We assume when PU is active it is engaged in a bidirectional communication with another PU, which is located far from $\rm SU_{tx}$ and hence its activity does not impact our analysis. We assume $\rm SU_{tx}$ is equipped with an M-beam RA (for spatial spectrum sensing, channel training and data transmission) with the capability of choosing one out of M sectors for its data transmission to $\rm SU_{rx}$, while $\rm SU_{rx}$ and PU use omni-directional antennas. We assume there is an error-free low-rate feedback channel[2] from $\rm SU_{rx}$ to $\rm SU_{tx}$, to enable $\rm SU_{tx}$ select the best sector for its data transmission to $\rm SU_{rx}$, and to adapt its transmit power according to the $\rm SU_{tx}$–$\rm SU_{rx}$ channel information. The direction (orientation) of PU and $\rm SU_{rx}$ with respect to $\rm SU_{tx}$ are denoted by angles $\phi_{\rm PU}$, and $\phi_{\rm SR}$, receptively, where $\phi_{\rm SR}, \phi_{\rm PU} \in (\phi_1, \phi_2)$. Clearly, in our problem $\rm SU_{tx}$ does not know these directions or angles (otherwise, the beam selection at $\rm SU_{tx}$ for data transmission would become trivial).

Let h, $h_{\rm ss}$, $h_{\rm sp}$ denote the fading coefficients of channels between $\rm SU_{tx}$ and PU, $\rm SU_{tx}$ and $\rm SU_{rx}$, and $\rm SU_{rx}$ and PU, respectively, when the RA of $\rm SU_{tx}$ is in omni-directional mode. We model these fading coefficients as independent zero mean circularly symmetric complex Gaussian random variables. Equivalently, $g = |h|^2$, $g_{\rm ss} = |h_{\rm ss}|^2$ and $g_{\rm sp} = |h_{\rm sp}|^2$ are independent exponentially distributed random variables

[1]Throughout this paper, "sector" and "beam" are used interchangeably.

[2]Given a low rate feedback, the error-free feedback channel is a reasonable assumption [18].

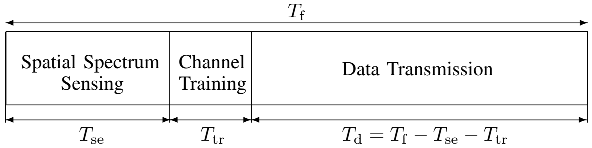


Fig. 2: The structure of frame employed by $\mathrm{SU_{tx}}$.

with mean γ, $\gamma_{\rm ss}$ and $\gamma_{\rm sp}$, respectively[3]. In our problem we assume that SUs and PU cannot cooperate, and hence SUs cannot estimate g and $g_{\rm sp}$. However, $\mathrm{SU_{tx}}$ knows the channel statistics, i.e., the mean values γ and $\gamma_{\rm sp}$. Let $\psi_{m'}$ and χ_m denote the fading coefficients of channel between m'-th sector of $\mathrm{SU_{tx}}$ and PU, and between m-th sector of $\mathrm{SU_{tx}}$ and $\mathrm{SU_{rx}}$, respectively, when the RA of $\mathrm{SU_{tx}}$ is in directional mode. Using the radiation pattern expression in (2) we can relate $\psi_{m'}$ to h and χ_m to $h_{\rm ss}$ as $\psi_{m'}=h\sqrt{p_{m'}(\phi_{\rm PU})}$, $\chi_m=h_{\rm ss}\sqrt{p_m(\phi_{\rm SR})}$. We assume the channel gain $\nu_m=|\chi_m|^2$ is an exponentially distributed random variable with mean α_m, and $\mathrm{SU_{tx}}$ knows α_m, for all m [32], [35]. For the readers' convenience, we have collected the most commonly used symbols in Table I.

TABLE I: Most commonly used symbols.

Symbol	Description
M	Number of beams
$N_{\rm se}$	Number of samples used for *spatial spectrum sensing*
$N_{\rm tr}$	Number of samples used for *channel training*
$P_{\rm tr}$	Power of training symbols
$\psi_{m'}$	Fading coefficient of channel between m'-th beam of $\mathrm{SU_{tx}}$ and PU
$\chi_m, \widehat{\chi}_m, \widetilde{\chi}_m$	Fading coefficient of channel between m-th beam of $\mathrm{SU_{tx}}$ and $\mathrm{SU_{rx}}$, LMMSE channel estimate, and its corresponding estimation error
$\alpha_m, \widehat{\alpha}_m, \widetilde{\alpha}_m$	Variances of $\chi_m, \widehat{\chi}_m, \widetilde{\chi}_m$
$m^*_{\rm PU}, m^*_{\rm SR}$	Indices of selected beam for PU and $\mathrm{SU_{rx}}$
$\widehat{\nu}^*$	Channel gain of selected beam for data transmission from $\mathrm{SU_{tx}}$ to $\mathrm{SU_{rx}}$

Suppose, SUs employ a frame with a fixed duration of $T_{\rm f}$ seconds, depicted in Fig. 2. We assume the $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ channel remains constant over the frame duration. $\mathrm{SU_{tx}}$ first senses the spectrum and monitors PU's activity. We refer to this period as *spatial spectrum sensing* phase with a variable duration of $T_{\rm se}=MN_{\rm se}T_{\rm s}$ seconds, where $T_{\rm s}$ is the sampling period and $N_{\rm se}$ is the number of collected samples during this phase per beam. Suppose $\mathcal{H}_1$ and $\mathcal{H}_0$ represent the binary hypotheses of PU being active and inactive, respectively, with prior probabilities $\Pr\{\mathcal{H}_1\}=\pi_1$ and $\Pr\{\mathcal{H}_0\}=\pi_0$. $\mathrm{SU_{tx}}$ applies a binary detection rule to decide whether or not PU is active. The details of the binary detector are presented in Section III-A. While being in this phase, $\mathrm{SU_{tx}}$ determines the beam corresponding to the orientation of PU based on the received signal energy as we describe in Section III-B.

Depending on the outcome of spectrum sensing, $\mathrm{SU_{tx}}$ stays in spatial spectrum sensing phase or enters the next phase, which we refer to as *channel training phase* with a variable duration of $T_{\rm tr}=MN_{\rm tr}T_{\rm s}$ seconds. In this phase, $\mathrm{SU_{tx}}$ sends $N_{\rm tr}$ training symbols with fixed symbol power $P_{\rm tr}$ per beam to enable channel estimation at $\mathrm{SU_{rx}}$, as we explain in Section IV-A. Based on the results of channel estimation for all beams, $\mathrm{SU_{rx}}$ selects the beam with the largest $\mathrm{SU_{tx}}$–$\mathrm{SU_{rx}}$ fading gain, as we describe in Section IV-B. This information as well as the corresponding beam index are shared with $\mathrm{SU_{tx}}$ via the feedback channel. Next, $\mathrm{SU_{tx}}$ enters *data transmission phase* with a variable duration of $T_{\rm d}=T_{\rm f}-T_{\rm se}-T_{\rm tr}$ seconds. During this phase, $\mathrm{SU_{tx}}$ sends $N_{\rm d}=T_{\rm d}/T_{\rm s}$ Gaussian data symbols with adaptive symbol power P to $\mathrm{SU_{rx}}$ over the selected strongest beam. $\mathrm{SU_{tx}}$ adapts P aiming at maximizing the achievable rates, subject to ATPC and AIC as we describe in Section V. In the following sections, we describe how $\mathrm{SU_{tx}}$ operates during spatial spectrum sensing phase, channel training phase, and data transmission phase.

[3]We note that the distances between users are included in the small scale fading model [34], i.e., the mean values $\gamma, \gamma_{\rm ss}, \gamma_{\rm sp}$ encompass distance-dependent path loss.

III. Spatial Spectrum Sensing Phase

A. Eigenvalue-Based Detector for Spatial Spectrum Sensing

Let $\widehat{\mathcal{H}}_1$ and $\widehat{\mathcal{H}}_0$ denote the detector outcome, i.e., the detector finds PU active (spectrum is sensed busy and occupied) and inactive (spectrum is sensed idle and unoccupied and thus can be used by $\mathrm{SU_{tx}}$ for data transmission), respectively. Suppose when PU is active, it transmits signal $s(t)$ with power $P_{\rm p}$. Let $y_m(n)$ denote the discrete-time representation of received signal at m-th sector of $\mathrm{SU_{tx}}$ at time instant $t=nT_{\rm s}$. We model PU's transmitted signal $s(n)$ as a zero-mean complex Gaussian random variable with variance $P_{\rm p}$ and we assume $\mathrm{SU_{tx}}$ knows $P_{\rm p}$. Since $\mathrm{SU_{tx}}$ collects $N_{\rm se}$ samples per beam during spatial spectrum sensing phase, the hypothesis testing problem at discrete time instant n for m-th sector is

$$\begin{aligned} \mathcal{H}_0:&\quad y_m(n)=w_m(n),\\ \mathcal{H}_1:&\quad y_m(n)=\psi_m(n)s(n)+w_m(n). \end{aligned} \tag{3}$$

The term $w_m(n)$ is the additive noise at m-th sector of $\mathrm{SU_{tx}}$ antenna and is modeled as $w_m(n)\sim\mathcal{CN}(0,\sigma_{\rm w}^2)$. We assume that $\psi_m(n)$, $s(n)$ and $w_m(n)$ are mutually independent random variables. Since $\mathrm{SU_{tx}}$ takes samples of the received signal for different sectors sequentially (in different time instants), $\psi_m(n)$ and $w_m(n)$ are independent and thus uncorrelated both in time and space (sector) domains. Under hypothesis $\mathcal{H}_1$, given ψ_m, we have $y_m(n)\sim\mathcal{CN}(0,\sigma_m^2+\sigma_{\rm w}^2)$ where $\sigma_m^2=|\psi_m|^2P_{\rm p}$. Under hypothesis $\mathcal{H}_0$, we have $y_m(n)\sim\mathcal{CN}(0,\sigma_{\rm w}^2)$.

Our proposed binary detector uses all the collected samples from M sectors. To facilitate the signal processing needed for the binary detection, we define an $M\times N_{\rm se}$ sample matrix $\boldsymbol{Z}=[\boldsymbol{z}_1,\dots,\boldsymbol{z}_{N_{\rm se}}]$, where the first row of $\boldsymbol{Z}$ is the $N_{\rm se}$ samples collected from the first sector, the second row of $\boldsymbol{Z}$ is the $N_{\rm se}$ samples collected from the second sector, and so forth. Given our assumptions, the columns of $\boldsymbol{Z}$ are orthogonal under both hypotheses, that is

$$\mathbb{E}\{\boldsymbol{z}_i\boldsymbol{z}_j^H|\mathcal{H}_0\}=\boldsymbol{0},\qquad \mathbb{E}\{\boldsymbol{z}_i\boldsymbol{z}_j^H|\mathcal{H}_1\}=\boldsymbol{0},\quad \text{for } i\neq j,\quad i,j=1,\dots,N_{\rm se} \tag{4}$$

where $\mathbb{E}\{\cdot\}$ is the statistical expectation operator and have the below covariance matrices

$$\boldsymbol{\Gamma}_0=\mathbb{E}\left\{\boldsymbol{z}_j\boldsymbol{z}_j^H|\mathcal{H}_0\right\}=\sigma_{\rm w}^2\boldsymbol{I}_M, \tag{5a}$$

$$\boldsymbol{\Gamma}_1=\mathbb{E}\left\{\boldsymbol{z}_j\boldsymbol{z}_j^H|\mathcal{H}_1,\boldsymbol{\psi}\right\}=P_{\rm p}\boldsymbol{\psi}\boldsymbol{\psi}^H+\sigma_{\rm w}^2\boldsymbol{I}_M, \tag{5b}$$

where vector $\boldsymbol{\psi} = \left[\psi_1, \psi_2, \ldots, \psi_M\right]^T$. Therefore the sample covariance matrix $\widehat{\boldsymbol{R}}$ becomes $\widehat{\boldsymbol{R}} = \frac{1}{N_{\rm se}}\boldsymbol{Z}\boldsymbol{Z}^H$. Let $f(\mathbf{Z}|\mathcal{H}_0)$ and $f(\mathbf{Z}|\mathcal{H}_1, \boldsymbol{\psi})$ denote the probability distribution function (pdf) of $\mathbf{Z}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ (given $\boldsymbol{\psi}$), respectively. These pdf expressions are

$$f(\mathbf{Z}|\mathcal{H}_0) = \frac{1}{(\pi\sigma_{\rm w}^2)^{N_{\rm eq}}} \exp\left\{\frac{\mathrm{tr}(\mathbf{Z}\mathbf{Z}^H)}{-\sigma_{\rm w}^2}\right\}, \tag{6a}$$

$$f(\mathbf{Z}|\mathcal{H}_1, \boldsymbol{\psi}) = \frac{1}{\pi^{N_{\rm eq}} \det(\boldsymbol{\Gamma_1})^{N_{\rm se}}} \exp\left\{\frac{\mathrm{tr}(\boldsymbol{\Gamma}_1^{-1}\mathbf{Z}\mathbf{Z}^H)}{-\sigma_{\rm w}^2}\right\}, \tag{6b}$$

where $N_{\rm eq} = MN_{\rm se}$. The optimal detector would compare the logarithm of likelihood ratio (LLR) against a threshold η_0 to detect the PU's activity as below

$$\mathrm{LLR} = \ln\frac{f(\mathbf{Z}|\mathcal{H}_1, \boldsymbol{\psi})}{f(\mathbf{Z}|\mathcal{H}_0)} \underset{\widehat{\mathcal{H}}_0}{\overset{\widehat{\mathcal{H}}_1}{\gtrless}} \eta_0. \tag{7}$$

In the absence of the knowledge of the fading coefficients vector $\boldsymbol{\psi}$, $\mathrm{SU_{tx}}$ obtains the generalized likelihood ratio test (GLRT) [36]–[40] which uses the maximum likelihood (ML) estimate of $\boldsymbol{\psi}$ under $\mathcal{H}_1$. Let $\mathcal{L}_1(\mathbf{Z}) = \ln f(\mathbf{Z}|\mathcal{H}_1, \boldsymbol{\psi})$. To find the maximum of $\mathcal{L}_1(\mathbf{Z})$ with respect to $\boldsymbol{\psi}$, we take the derivative of $\mathcal{L}_1(\mathbf{Z})$ with respect to $\boldsymbol{\psi}$ and solve $\frac{\partial}{\partial\boldsymbol{\psi}}\mathcal{L}_1(\mathbf{Z}) = \mathbf{0}$ for $\boldsymbol{\psi}$. The obtained solution is the ML estimate of $\boldsymbol{\psi}$. Substituting this solution into (7) and after some mathematical manipulation, we reach the following decision rule $T = \frac{\lambda_{\rm max}}{\sigma_{\rm w}^2} \underset{\widehat{\mathcal{H}}_0}{\overset{\widehat{\mathcal{H}}_1}{\gtrless}} \eta$ [39], where T is the test statistics, $\lambda_{\rm max}$ is the maximum eigenvalue of $\widehat{\boldsymbol{R}}$, and η is the threshold. For large $N_{\rm se}$, T under $\mathcal{H}_0$ is distributed as Tracy-Widom distribution of order 2 [39, Lemma 1] and the probability of false alarm $P_{\rm fa} = \Pr(\widehat{\mathcal{H}}_1|\mathcal{H}_0) = \Pr(T > \eta|\mathcal{H}_0)$ is

$$P_{\rm fa} = 1 - F_{\rm TW2}\left(\frac{\eta - \theta_{\rm sen}}{\sigma_{\rm sen}}\right), \tag{8}$$

where $F_{\rm TW2}(\cdot)$ is the commutative distribution function (CDF) of Tracy-Widom distribution of order 2 and $\theta_{\rm sen}$ and $\sigma_{\rm sen}$ in (8) are given below

$$\theta_{\rm sen} = \left(1 + \sqrt{\frac{M}{N_{\rm se}}}\right)^2, \tag{9a}$$

$$\sigma_{\rm sen} = \frac{1}{\sqrt{N_{\rm se}}}\left(1 + \sqrt{\frac{M}{N_{\rm se}}}\right)\left(\frac{1}{\sqrt{N_{\rm se}}} + \frac{1}{\sqrt{M}}\right)^{\frac{1}{3}}. \tag{9b}$$

For large $N_{\rm se}$, T under $\mathcal{H}_1$ is Gaussian distributed [39, Lemma 2] and the probability of detection $P_{\rm d} = \Pr(\widehat{\mathcal{H}}_1|\mathcal{H}_1) = \Pr(T > \eta|\mathcal{H}_1)$ is [39], [40]

$$P_{\rm d} = Q\left(\frac{\eta\sqrt{N_{\rm se}}}{1 + \delta_{\rm sen}} - \frac{M - 1}{\delta_{\rm sen}\sqrt{N_{\rm se}}} - \sqrt{N_{\rm se}}\right), \tag{10}$$

where $\delta_{\rm sen} = \frac{P_{\rm p}\|\boldsymbol{\psi}\|^2}{\sigma_{\rm w}^2}$. The average detection probability $\overline{P}_{\rm d}$ can be computed by averaging (10) over vector $\boldsymbol{\psi}$, $\overline{P}_{\rm d} = \mathbb{E}_{\boldsymbol{\psi}}\{P_{\rm d}\}$. For a given $\overline{P}_{\rm d}$, we can numerically find η and obtain $\overline{P}_{\rm fa}$ using (8). We can also compute the probabilities of events $\widehat{\mathcal{H}}_0$ and $\widehat{\mathcal{H}}_1$ as $\widehat{\pi}_0 = \Pr\{\widehat{\mathcal{H}}_0\} = \beta_0 + \beta_1$ and $\widehat{\pi}_1 = \Pr\{\widehat{\mathcal{H}}_1\} = 1 - \widehat{\pi}_0$, respectively, where

$$\beta_0 = \Pr\{\mathcal{H}_0, \widehat{\mathcal{H}}_0\} = \pi_0(1 - \overline{P}_{\rm fa}), \tag{11a}$$

$$\beta_1 = \Pr\{\mathcal{H}_1, \widehat{\mathcal{H}}_0\} = \pi_1(1 - \overline{P}_{\rm d}). \tag{11b}$$

B. Determining the Beam Corresponding to PU Direction

During spatial spectrum sensing phase when the spectrum is sensed busy, $\mathrm{SU_{tx}}$ determines the beam corresponding to the direction of PU based on the received signal energy. Let ε_m be the energy of received signal at m-th beam. We have

$$\varepsilon_m = \frac{1}{N_{\rm se}} \sum_{n=1+(m-1)N_{\rm se}}^{mN_{\rm se}} \left|y_m(n)\right|^2. \tag{12}$$

$\mathrm{SU_{tx}}$ determines the beam with the largest amount of received energy $m^*_{\rm PU} = \arg\max\{\varepsilon_m\}$ among all beams. For large $N_{\rm se}$, we invoke central limit theorem (CLT) to approximate ε_m's as Gaussian random variables under both hypotheses. Thus, under $\mathcal{H}_0$ we approximate ε_m as a Gaussian with distribution $\varepsilon_m \sim \mathcal{N}(\sigma_{\rm w}^2, \sigma_{\rm w}^4/N_{\rm se})$. Similarly, under $\mathcal{H}_1$, given $\phi_{\rm PU}$ we approximate ε_m as another Gaussian with distribution $\varepsilon_m \sim \mathcal{N}(\varrho_m, \sigma^2_{\varepsilon_m|\mathcal{H}_1})$, where the mean $\varrho_m = \gamma P_{\rm p}\, p_m(\phi_{\rm PU}) + \sigma_{\rm w}^2$, and the variance $\sigma^2_{\varepsilon_m|\mathcal{H}_1}$ is given below

$$\sigma^2_{\varepsilon_m|\mathcal{H}_1} = \frac{1}{N_{\rm se}}\Big[\sigma_{\rm w}^4 + 3P_{\rm p}^2\gamma^2\, p_m^2(\phi_{\rm PU}) + 2\sigma_{\rm w}^2 P_{\rm p}\gamma\, p_m(\phi_{\rm PU})\Big]. \tag{13}$$

We note that, there is a non-zero error probability when $\mathrm{SU_{tx}}$ determines the beam index $m^*_{\rm PU}$, i.e., it is possible that $m^*_{\rm PU}$ is not the true beam index corresponding to PU direction.

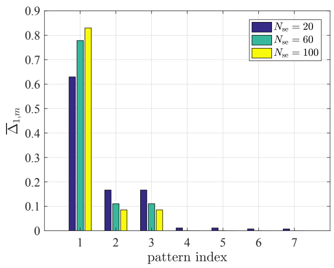
Let $\overline{\Delta}_{i,m}$ represent the average error probability of finding the sector index corresponding to PU direction, i.e., the probability that $m^*_{\rm PU} = i$ while the true PU direction lies in the angular domain of m-th sector, $\phi_{\rm PU} \in \Phi_m = \left[\frac{2\pi(m-3/2)}{M}, \frac{2\pi(m-1/2)}{M}\right)$, for $i \neq m, i, m = 1, \ldots, M$. To find $\overline{\Delta}_{i,m}$ we start with finding $\Delta_i = \Pr\{m^*_{\rm PU} = i|\phi_{\rm PU}, \widehat{\mathcal{H}}_1\}$, which is the probability that the index of selected sector is i, given $\phi_{\rm PU}$ and $\widehat{\mathcal{H}}_1$ (the binary detector in Section III-A finds PU active). Note that under both hypotheses, ε_m's are independent. Also, under $\mathcal{H}_0$, ε_m's are identically distributed. Therefore, we have

$$\begin{aligned}\Delta_i &= \Pr\Big\{\underset{\forall m,\, m\neq i}{\varepsilon_i > \varepsilon_m}\,\Big|\phi_{\rm PU}, \widehat{\mathcal{H}}_1\Big\} \\ &= \varsigma_1 \int_0^\infty f_{\varepsilon_i|\mathcal{H}_1}(y|\phi_{\rm PU}) \prod_{\substack{m=1\\ m\neq i}}^{M} F_{\varepsilon_m|\mathcal{H}_1}(y|\phi_{\rm PU})dy \\ &\quad + \varsigma_0 \int_0^\infty f_{\varepsilon_m|\mathcal{H}_0}(y) F^{M-1}_{\varepsilon_m|\mathcal{H}_0}(y)dy\end{aligned} \tag{14}$$

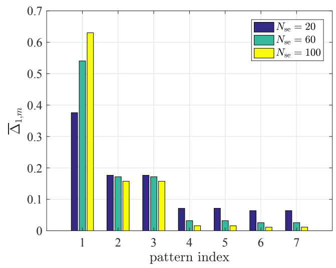
where $f_{\varepsilon_m|\mathcal{H}_\ell}(x)$ and $F_{\varepsilon_m|\mathcal{H}_\ell}(x)$ are the pdf and CDF expressions of ε_m under $\mathcal{H}_\ell, \ell = 0, 1$ and

$$\varsigma_0 = \Pr\{\mathcal{H}_0|\widehat{\mathcal{H}}_1\} = \frac{\pi_0\overline{P}_{\rm fa}}{\widehat{\pi}_1}, \tag{15a}$$

$$\varsigma_1 = \Pr\{\mathcal{H}_1|\widehat{\mathcal{H}}_1\} = \frac{\pi_1\overline{P}_{\rm d}}{\widehat{\pi}_1}. \tag{15b}$$



(a)



(b)

Fig. 3: $\overline{\Delta}_{1,m}$ versus the index beam m for $\phi_{3\text{dB}}=20^{\circ}$ (a) $\text{SNR}_{\text{PU}}=0\,\text{dB}$, (b) $\text{SNR}_{\text{PU}}=-5\,\text{dB}$.

Using Δ_i, we find $\overline{\Delta}_{i,m}$ as the following

$$\overline{\Delta}_{i,m} = \int_{\phi_{\text{PU}}\in\Phi_m} \Delta_i \Pr\{\phi_{\text{PU}}\in\Phi_m\}\, d\phi_{\text{PU}}. \tag{16}$$

Note that $\overline{\Delta}_{i,i}$ is the probability of selecting the correct beam and $\overline{\Delta}_{i,m}$ for $i\neq m$ is the probability of selecting the incorrect beam, leading to error probability in beam selection. The average error probability $\overline{\Delta}_{1,m}$ versus the index beam m is shown in Figs. 3a and 3b for $\text{SNR}_{\text{PU}}=\gamma P_{\text{p}}/\sigma^2_{\text{w}}=0,-5\,\text{dB}$. As expected, $\overline{\Delta}_{1,1}$ increases and $\overline{\Delta}_{1,m}, m\neq 1$ decreases as N_{se} increases.

IV. Channel Training Phase

A. Channel Estimation at SU$_{\text{rx}}$

During this phase, SU$_{\text{tx}}$ sends the training vector $\boldsymbol{x}_t$ over all beams to enable channel estimation at SU$_{\text{rx}}$. Without loss of generality, we assume $\boldsymbol{x}_t = \sqrt{P_{\text{tr}}}\,\mathbf{1}$, where $\mathbf{1}$ is an $N_{\text{tr}}\times 1$ all-ones vector and P_{tr} is given. Let $\boldsymbol{r}_m = \big[r_m(1),\ldots,r_m(N_{\text{tr}})\big]^T$ denote the discrete-time representation of received training symbols at SU$_{\text{rx}}$ from m-th sector of SU$_{\text{tx}}$. We note that SU$_{\text{tx}}$ enters this phase when the outcome of the binary detector in Section III-A is $\widehat{\mathcal{H}}_0$. Due to error in spatial spectrum sensing, we need to differentiate the signal model for r_m under $\mathcal{H}_0$ and $\mathcal{H}_1$. Assuming the fading coefficient χ_m is unchanged during the frame, we have

$$\begin{aligned}\mathcal{H}_0,\widehat{\mathcal{H}}_0:&\quad r_m(n)=\chi_m\sqrt{P_{\text{tr}}}+q_m(n),\\ \mathcal{H}_1,\widehat{\mathcal{H}}_0:&\quad r_m(n)=\chi_m\sqrt{P_{\text{tr}}}+h_{\text{sp}}(n)\,s(n)+q_m(n),\end{aligned} \tag{17}$$

where $q_m(n)$ is the additive noise at SU$_{\text{rx}}$ antenna and is modeled as $q_m(n)\sim\mathcal{CN}\big(0,\sigma^2_{\text{q}}\big)$. The linear minimum mean square error (LMMSE) estimation of fading coefficient χ_m when the spectrum sensing result is $\widehat{\mathcal{H}}_0$ can be obtained as [41]

$$\widehat{\chi}_m = C_{\chi_m\boldsymbol{r}_m} C^{-1}_{\boldsymbol{r}_m}\boldsymbol{r}_m, \tag{18a}$$

$$C_{\chi_m\boldsymbol{r}_m} = \mathbb{E}\{\chi_m\boldsymbol{r}^H_m|\widehat{\mathcal{H}}_0\} = \sqrt{P_{\text{tr}}}\,\alpha_m\,\mathbf{1}, \tag{18b}$$

$$\begin{aligned}C_{\boldsymbol{r}_m} &= \mathbb{E}\big\{\boldsymbol{r}_m\boldsymbol{r}^H_m|\widehat{\mathcal{H}}_0\big\} = \omega_0\,\mathbb{E}\big\{\boldsymbol{r}_m\boldsymbol{r}^H_m|\mathcal{H}_0,\widehat{\mathcal{H}}_0\big\}\\ &\quad+\omega_1\,\mathbb{E}\big\{\boldsymbol{r}_m\boldsymbol{r}^H_m|\mathcal{H}_1,\widehat{\mathcal{H}}_0\big\},\end{aligned} \tag{18c}$$

where

$$\omega_0 = \Pr\{\mathcal{H}_0|\widehat{\mathcal{H}}_0\} = \frac{\pi_0(1-\overline{P}_{\text{fa}})}{\widehat{\pi}_0} = \frac{\beta_0}{\widehat{\pi}_0}, \tag{19a}$$

$$\omega_1 = \Pr\{\mathcal{H}_1|\widehat{\mathcal{H}}_0\} = \frac{\pi_1(1-\overline{P}_{\text{d}})}{\widehat{\pi}_0} = \frac{\beta_1}{\widehat{\pi}_0}. \tag{19b}$$

Finally, the LMMSE estimation of χ_m when the spectrum sensing result is $\widehat{\mathcal{H}}_0$, given in (18a), reduces to

$$\widehat{\chi}_m = \frac{\alpha_m\sqrt{P_{\text{tr}}}}{\alpha_m P_{\text{tr}} N_{\text{tr}}+\sigma^2_{\text{q}}+\omega_1\sigma^2_{\text{p}}}\sum_{n=1}^{N_{\text{tr}}} r_m(n), \tag{20}$$

where $\sigma^2_{\text{p}} = P_{\text{p}}\gamma_{\text{sp}}$. The estimation error is $\widetilde{\chi}_m = \chi_m - \widehat{\chi}_m$ where $\widehat{\chi}_m$ and $\widetilde{\chi}_m$ are orthogonal random variables [41], and $\widehat{\chi}_m$ and $\widetilde{\chi}_m$ are zero mean. Approximating $h_{\text{sp}}(n)s(n)$ as a zero-mean Gaussian random variable with variance σ^2_{p}, we find that the estimate $\widehat{\chi}_m$ is distributed as a Gaussian mixture random variable [16], [42]. Let $\widehat{\alpha}_m$ and $\widetilde{\alpha}_m$, represent the variances of $\widehat{\chi}_m$ and $\widetilde{\chi}_m$, respectively. Also, Let $\widehat{\alpha}^0_m$ and $\widehat{\alpha}^1_m$ represent the variances of $\widehat{\chi}_m$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. We have

$$\widehat{\alpha}^0_m = \mathbb{VAR}\{\widehat{\chi}_m|\mathcal{H}_0,\widehat{\mathcal{H}}_0\} = \frac{\alpha^2_m P_{\text{tr}} N_{\text{tr}}\big(\alpha_m P_{\text{tr}} N_{\text{tr}}+\sigma^2_{\text{q}}\big)}{\big(\alpha_m P_{\text{tr}} N_{\text{tr}}+\sigma^2_{\text{q}}+\omega_1\sigma^2_{\text{p}}\big)^2}, \tag{21a}$$

$$\widehat{\alpha}^1_m = \mathbb{VAR}\{\widehat{\chi}_m|\mathcal{H}_1,\widehat{\mathcal{H}}_0\} = \frac{\alpha^2_m P_{\text{tr}} N_{\text{tr}}\big(\alpha_m P_{\text{tr}} N_{\text{tr}}+\sigma^2_{\text{q}}+\sigma^2_{\text{p}}\big)}{\big(\alpha_m P_{\text{tr}} N_{\text{tr}}+\sigma^2_{\text{q}}+\omega_1\sigma^2_{\text{p}}\big)^2}. \tag{21b}$$

Therefore, $\widehat{\alpha}_m = \omega_0\,\widehat{\alpha}^0_m + \omega_1\,\widehat{\alpha}^1_m$. Also, let $\widetilde{\alpha}^0_m$ and $\widetilde{\alpha}^1_m$ indicate the variances of $\widetilde{\chi}_m$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. We have

$$\widetilde{\alpha}^0_m = \mathbb{VAR}\{\widetilde{\chi}_m|\mathcal{H}_0,\widehat{\mathcal{H}}_0\} = \alpha_m - \widehat{\alpha}^0_m, \tag{22a}$$

$$\widetilde{\alpha}^1_m = \mathbb{VAR}\{\widetilde{\chi}_m|\mathcal{H}_1,\widehat{\mathcal{H}}_0\} = \alpha_m - \widehat{\alpha}^1_m. \tag{22b}$$

Hence, $\widetilde{\alpha}_m = \omega_0\,\widetilde{\alpha}^0_m + \omega_1\,\widetilde{\alpha}^1_m$. For perfect spectrum sensing, we get $\omega_0 = 1$ and $\omega_1 = 0$ and $\widehat{\chi}_m$ becomes Gaussian.

B. Determining the Beam Corresponding to SU$_{\text{rx}}$ Direction

SU$_{\text{rx}}$ finds $\widehat{\chi}_m$ for all beams Consider the random variable $\widehat{\nu}_m = |\widehat{\chi}_m|^2$. Under hypothesis $\mathcal{H}_\ell, \ell=0,1$, given $\widehat{\mathcal{H}}_0$, $\widehat{\nu}_m$ is an exponential random variable with mean $\widehat{\alpha}^\ell_m$ and pdf

$$f^\ell_{\widehat{\nu}_m}(y) = \frac{1}{\widehat{\alpha}^\ell_m} e^{\frac{-y}{\widehat{\alpha}^\ell_m}}. \tag{23}$$

Hence, the pdf of $\widehat{\nu}_m$ can be written as

$$f_{\widehat{\nu}_m}(y) = \omega_0\, f^0_{\widehat{\nu}_m}(y) + \omega_1\, f^1_{\widehat{\nu}_m}(y). \tag{24}$$

SU$_{\text{rx}}$ obtains $\widehat{\nu}^* = \max\{\widehat{\nu}_m\}$ among all beams and the corresponding beam index $m^*_{\text{SR}} = \arg\max\{\widehat{\nu}_m\}$ and feeds back this information to SU$_{\text{tx}}$. Let $\Psi^\ell_i = \Pr\{m^*_{\text{SR}} = i|\mathcal{H}_\ell,\widehat{\mathcal{H}}_0\}$ denote the probability that $m^*_{\text{SR}} = i$ under hypothesis $\mathcal{H}_\ell$ and the binary detector outcome is $\widehat{\mathcal{H}}_0$. To characterize Ψ^ℓ_i we need to find the CDF and pdf of $\widehat{\nu}^*$ given $\mathcal{H}_\ell$, denoted as $F^\ell_{\widehat{\nu}^*}(\cdot)$ and $f^\ell_{\widehat{\nu}^*}(\cdot)$, respectively. Note that given our assumptions, $\widehat{\nu}_m$'s are independent across sectors, however, not necessarily identically distributed. Therefore, the CDF $F^\ell_{\widehat{\nu}^*}(x)$ can be written as

$$F^{\ell}_{\widehat{\nu}^*}(y)=\prod_{m=1}^{M} F^{\ell}_{\widehat{\nu}_m}(y)=1+\sum_{m=1}^{M}(-1)^m \sum\nolimits_m e^{-yA^{\ell}_{j_1:j_m}} \quad (25)$$

$$A^{\ell}_{j_1:j_m}=\sum_{i=1}^{m}\frac{1}{\widehat{\alpha}^{\ell}_{j_i}}, \qquad \sum\nolimits_m=\sum_{j_1=1}^{M-m+1}\sum_{j_2=j_1+1}^{M-m+2}\cdots\sum_{j_m=j_{m-1}+1}^{M}.$$

From the CDF in (25), we can find the pdf $f^{\ell}_{\widehat{\nu}^*}(y)$

$$\begin{aligned} f^{\ell}_{\widehat{\nu}^*}(y) &= \sum_{i=1}^{M} f^{\ell}_{\widehat{\nu}_i}(y) \prod_{\substack{m=1\\ m\neq i}}^{M} F^{\ell}_{\widehat{\nu}_m}(y) \\ &= \sum_{m=1}^{M}(-1)^{m+1}\sum\nolimits_m A^{\ell}_{j_1:j_m} e^{-yA^{\ell}_{j_1:j_m}}. \end{aligned} \quad (26)$$

Similar to section III-B, we obtain Ψ^{ℓ}_i as

$$\Psi^{\ell}_i=\int_0^{\infty} f^{\ell}_{\widehat{\nu}_i}(y)\prod_{\substack{m=1\\ m\neq i}}^{M} F^{\ell}_{\widehat{\nu}_m}(y)\, dy. \quad (27)$$

Without loss of generality, suppose $i=1$. After some mathematical simplification, Ψ^{ℓ}_1 can be expressed as

$$\Psi^{\ell}_1 = 1+\sum_{m=1}^{M-1}(-1)^m \sum\nolimits'_m \frac{1}{1+\widehat{\alpha}^{\ell}_1 B^{\ell}_{j_1:j_m}}, \quad (28)$$

where

$$B^{\ell}_{j_1:j_m}=\sum_{i=1}^{m}\frac{1}{\widehat{\alpha}^{\ell}_{1+j_i}}, \qquad \sum\nolimits'_m=\sum_{j_1=1}^{M-m}\sum_{j_2=j_1+1}^{M-m+1}\cdots\sum_{j_m=j_{m-1}+1}^{M-1}.$$

Then, we have $\Psi_i=\Pr\{m^*_{\rm SR}=i|\widehat{\mathcal{H}}_0\}=\omega_0\,\Psi^0_i+\omega_1\,\Psi^1_i$.

V. Data Transmission Phase

During this phase, $\mathrm{SU_{tx}}$ sends Gaussian data symbols to $\mathrm{SU_{rx}}$, while data symbol transmission power is adapted based on the information provided by $\mathrm{SU_{rx}}$ through the feedback channel. In particular, $\mathrm{SU_{tx}}$ transmits $x(n)\sim\mathcal{CN}(0,P)$ over the selected beam $i=m^*_{\rm SR}$, where P depends on $\widehat{\nu}_i$, and symbols are independent and identically distributed (i.i.d). Let $u(n)$ denote the discrete-time representation of received signal at $\mathrm{SU_{rx}}$ from i-th beam of $\mathrm{SU_{tx}}$. We note that $\mathrm{SU_{tx}}$ enters this phase when the outcome of the binary detector in Section III-A is $\widehat{\mathcal{H}}_0$. Due to error in spatial spectrum sensing, we need to distinguish the signal model for $u(n)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$. We have

$$\begin{aligned} &\mathcal{H}_0,\widehat{\mathcal{H}}_0: \quad u(n)=\chi_i\,x(n)+q(n),\\ &\mathcal{H}_1,\widehat{\mathcal{H}}_0: \quad u(n)=\chi_i\,x(n)+h_{\rm sp}(n)\,s(n)+q(n), \end{aligned} \quad (29)$$

where $q(n)\sim\mathcal{CN}(0,\sigma^2_{\rm q})$ and are i.i.d. Substituting $\chi_i=\widehat{\chi}_i+\widetilde{\chi}_i$ in (29), we reach at

$$\begin{aligned} &\mathcal{H}_0,\widehat{\mathcal{H}}_0: \quad u(n)=\widehat{\chi}_i\,x(n)+\overbrace{\widetilde{\chi}_i\,x(n)+q(n)}^{\text{new noise }\eta_{i,0}(n)},\\ &\mathcal{H}_1,\widehat{\mathcal{H}}_0: \quad u(n)=\widehat{\chi}_i\,x(n)+\underbrace{\widetilde{\chi}_i\,x(n)+h_{\rm sp}(n)s(n)+q(n)}_{\text{new noise }\eta_{i,1}(n)}. \end{aligned} \quad (30)$$

We obtain an achievable rate expression for a frame by considering symbol-wise mutual information between channel input and output over the duration of $N_{\rm d}$ data symbols as follows

$$\begin{aligned} R &= \frac{D_{\rm d}}{N_{\rm d}}\sum_{n=1}^{N_{\rm d}}\mathbb{E}\left\{I\left(x(n);u(n)\big|\widehat{\boldsymbol{\nu}},\widehat{\mathcal{H}}_0\right)\right\}\\ &= \frac{D_{\rm d}}{N_{\rm d}}\sum_{n=1}^{N_{\rm d}}\Big[\beta_0\,\mathbb{E}\left\{I\left(x(n);u(n)\big|\widehat{\boldsymbol{\nu}},\mathcal{H}_0,\widehat{\mathcal{H}}_0\right)\right\}\\ &\qquad\qquad + \beta_1\,\mathbb{E}\left\{I\left(x(n);u(n)\big|\widehat{\boldsymbol{\nu}},\mathcal{H}_1,\widehat{\mathcal{H}}_0\right)\right\}\Big], \end{aligned} \quad (31)$$

where $D_{\rm d}=T_{\rm d}/T_{\rm f}$ is the fraction of the frame used for data transmission and the expectations are taken over $\widehat{\boldsymbol{\nu}}=[\widehat{\nu}_1,\dots,\widehat{\nu}_M]$ given $\widehat{\mathcal{H}}_0$ and $\mathcal{H}_\ell, \ell=0,1$. To characterize R in (31) we need to find $\mathbb{E}\{I(x(n);u(n)|\widehat{\boldsymbol{\nu}},\mathcal{H}_\ell,\widehat{\mathcal{H}}_0)\}$ which is given in (32). Term 1 in (32) is the mutual information between $x(n)$ and $u(n)$ when $\mathrm{SU_{tx}}$ transmits over j-th beam, given the estimated channel gain $\widehat{\nu}_j=|\widehat{\chi}_j|^2$, and given $\mathcal{H}_\ell$ and $\widehat{\mathcal{H}}_0$. Term 2 in (32) is the pdf of estimated channel gain $\widehat{\nu}_j=|\widehat{\chi}_j|^2$ when j-th beam is the selected strongest beam, and is characterized by statistics of channel estimation error and beam selection error, occurred during channel training phase. Focusing on Term 1 in (32) we have

$$\begin{aligned} I\left(x(n);u(n)\big|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\right) &= h\big(x(n)\big|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\big)\\ &\quad - h\big(x(n)\big|u(n),\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\big), \end{aligned} \quad (33)$$

where $h(\cdot)$ is the differential entropy. From now on, we drop the variable n in $x(n)$ and $u(n)$ for brevity. Consider the first term in (33). Since $x\sim\mathcal{CN}(0,P)$ we have $h(x|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell)=\log_2(\pi eP)$. Consider the second term in (33). Due to channel estimation error, the new noises $\eta_{i,\ell}$ in (30) are non-Gaussian and this term does not have a closed form expression. Hence, similar to [43]–[45] we employ bounding techniques to find an upper bound on this term. This term is upper bounded by the entropy of a Gaussian random variable with the variance $\Theta^{i,\ell}_{\rm M}$

$$\Theta^{i,\ell}_{\rm M}=\mathbb{E}\left\{\left|x-\mathbb{E}\{x\,|\,\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\}\right|^2\right\}, \quad (34)$$

where the expectations are taken over the conditional pdf of x given $u,\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell$. In fact, $\Theta^{i,\ell}_{\rm M}$ is the mean square error (MSE) of the MMSE estimate of x given $u,\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell$. Using minimum variance property of MMSE estimator, we have $\Theta^{i,\ell}_{\rm M}\leq\Theta^{i,\ell}_{\rm L}$, where $\Theta^{i,\ell}_{\rm L}$ is the MSE of the LMMSE estimate of x given $u,\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell$. Combining all, we find $h(x|u,\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell)\leq\log_2(\pi e\Theta^{i,\ell}_{\rm L})$ and $I(x,u|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell)\geq\log_2(P/\Theta^{i,\ell}_{\rm L})$ where

$$\Theta^{i,\ell}_{\rm L}=\frac{P\sigma^2_{\eta_{i,\ell}}}{\sigma^2_{\eta_{i,\ell}}+\widehat{\nu}_iP}, \qquad \sigma^2_{\eta_{i,\ell}}=\widetilde{\alpha}^{\ell}_iP+\sigma^2_{\rm q}+\ell\sigma^2_{\rm p}. \quad (35)$$

At the end, we obtain the lower bounds as follow

$$I\left(x;u|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_0\right)\geq\log_2\Big(1+\frac{\widehat{\nu}_iP}{\widetilde{\alpha}^0_iP+\sigma^2_{\rm q}}\Big), \quad (36a)$$

$$I\left(x;u|\widehat{\nu}_i,\widehat{\mathcal{H}}_0,\mathcal{H}_1\right)\geq\log_2\Big(1+\frac{\widehat{\nu}_iP}{\widetilde{\alpha}^1_iP+\sigma^2_{\rm q}+\sigma^2_{\rm p}}\Big). \quad (36b)$$

$$\mathbb{E}\Big\{I\left(x(n);u(n)|\widehat{\boldsymbol{\nu}},\mathcal{H}_\ell,\widehat{\mathcal{H}}_0\right)\Big\} = \int_{\widehat{\nu}_1=0}^{\infty} I\left(x(n);u(n)|\widehat{\nu}_1,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\right) f_{\widehat{\nu}_1}^{\ell}(\widehat{\nu}_1)\Pr\left(v_1>v_m \text{ for } m=2,...,M|\mathcal{H}_\ell,\widehat{\mathcal{H}}_0\right) d\widehat{\nu}_1$$
$$+\dots$$
$$+\int_{\widehat{\nu}_M=0}^{\infty} I\left(x(n);u(n)|\widehat{\nu}_M,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\right) f_{\widehat{\nu}_M}^{\ell}(\widehat{\nu}_M)\Pr\left(v_M>v_m \text{ for } m=1,...,M-1|\mathcal{H}_\ell,\widehat{\mathcal{H}}_0\right) d\widehat{\nu}_M$$
$$=\sum_{j=1}^{M}\int_{\widehat{\nu}_j=0}^{\infty} \underbrace{I\left(x(n);u(n)|\widehat{\nu}_j,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell\right)}_{\text{Term 1}} \underbrace{f_{\widehat{\nu}_j}^{\ell}(\widehat{\nu}_j)\prod_{\substack{m=1\\m\neq j}}^{M} F_{\widehat{\nu}_m}^{\ell}(\widehat{\nu}_j)}_{\text{Term 2}}\, d\widehat{\nu}_j. \tag{32}$$

Substituting equations (32) and (36) in (31) and changing the integration variable (replacing $\widehat{\nu}_j$ with y), we reach at

$$R \geq R_{\text{LB}} = D_{\text{d}}\beta_0 R_0 + D_{\text{d}}\beta_1 R_1 \tag{37}$$

where

$$R_0=\sum_{j=1}^{M}\int_0^{\infty}\log_2\Big(1+\frac{yP}{\widetilde{\alpha}_j^0P+\sigma_{\text{q}}^2}\Big)f_{\widehat{\nu}_j}^0(y)\prod_{\substack{m=1\\m\neq j}}^{M}F_{\widehat{\nu}_m}^0(y)dy,$$

$$R_1=\sum_{j=1}^{M}\int_0^{\infty}\log_2\Big(1+\frac{yP}{\widetilde{\alpha}_j^1P+\sigma_{\text{q}}^2+\sigma_{\text{p}}^2}\Big)f_{\widehat{\nu}_j}^1(y)\prod_{\substack{m=1\\m\neq j}}^{M}F_{\widehat{\nu}_m}^1(y)dy.$$

We note that the lower bounds in (36) are achieved when the new noises $\eta_{m,0},\eta_{m,1}$ in (30) are regarded as worst-case Gaussian noise and hence the MMSE and LMMSE of x given $u,\widehat{\nu}_m,\widehat{\mathcal{H}}_0,\mathcal{H}_\ell$ coincide.

So far, we have established a lower bound on the achievable rates. Next, we characterize AIC and ATPC. Let $\overline{I}_{\text{av}}$ indicate the maximum allowed interference power imposed on PU. To satisfy the AIC, we need to have

$$\beta_1\,\mathbb{E}\{g\}\Big[D_{\text{d}}\mathbb{E}\{p(\kappa_{\text{SR}}^*-\kappa_{\text{PU}}^*)\,P\,|\,\mathcal{H}_1,\widehat{\mathcal{H}}_0\} \tag{38}$$
$$+D_{\text{tr}}P_{\text{tr}}\sum_{j=1}^{M}\mathbb{E}\big\{p(\kappa_j-\kappa_{\text{PU}}^*)\,|\,\mathcal{H}_1,\widehat{\mathcal{H}}_0\big\}\Big]\leq\overline{I}_{\text{av}},$$

where $D_{\text{tr}}=T_{\text{tr}}/T_{\text{f}}$. The first term in (38) is the average interference imposed on PU when SU_{tx} transmits data symbols, and the second term is the average interference imposed on PU when SU_{tx} sends training symbols for channel estimation at SU_{rx}. Consider the two conditional expectation terms inside the bracket in (38). Using the fact that, given $\mathcal{H}_1,\widehat{\mathcal{H}}_0$, $p(\cdot)$ and P (which depends on $\widehat{\nu}^*$) are independent, and also the average probabilities derived in (16) and (27) we have

$$\mathbb{E}\{p(\kappa_{\text{SR}}^*-\kappa_{\text{PU}}^*)|\mathcal{H}_1,\widehat{\mathcal{H}}_0\}=\sum_{j=1}^{M}\sum_{i=1}^{M}\Psi_j^1\,\overline{\Delta}_{m_{\text{PU}}^*,i}\;p(\kappa_j-\kappa_i), \tag{39}$$

$$\mathbb{E}\{p(\kappa_j-\kappa_{\text{PU}}^*)\,|\,\mathcal{H}_1,\widehat{\mathcal{H}}_0\}=\sum_{i=1}^{M}\overline{\Delta}_{m_{\text{PU}}^*,i}\;p(\kappa_j-\kappa_i). \tag{40}$$

Then, the constraint in (38) can be written as

$$D_{\text{d}}\,b_0\,\mathbb{E}\big\{P|\mathcal{H}_1,\widehat{\mathcal{H}}_0\big\}+D_{\text{tr}}u_0P_{\text{tr}}\leq\overline{I}_{\text{av}}, \tag{41}$$

where

$$b_0=\beta_1\gamma\sum_{j=1}^{M}\sum_{i=1}^{M}\Psi_j^1\,\overline{\Delta}_{m_{\text{PU}}^*,i}\;p(\kappa_j-\kappa_i), \tag{42a}$$

$$u_0=\beta_1\gamma\sum_{j=1}^{M}\sum_{i=1}^{M}\overline{\Delta}_{m_{\text{PU}}^*,i}\;p(\kappa_j-\kappa_i), \tag{42b}$$

$$\mathbb{E}\big\{P|\mathcal{H}_1,\widehat{\mathcal{H}}_0\big\}=\int_0^{\infty}P(y)f_{\widehat{\nu}^*}^1(y)dy. \tag{42c}$$

We note that spectrum sensing error, PU beam selection error, and SU_{rx} beam selection error are reflected in AIC through variables β_1, $\overline{\Delta}_{m_{\text{PU}}^*,i}$ and Ψ_j^1, respectively. Also, channel estimation error influences AIC through variable P. Let $\overline{P}_{\text{av}}$ denote the maximum allowed average transmit power of SU_{tx}. To satisfy the ATPC, we need to have

$$\beta_0D_{\text{d}}\mathbb{E}\big\{P|\mathcal{H}_0,\widehat{\mathcal{H}}_0\big\}+\beta_1D_{\text{d}}\mathbb{E}\big\{P|\mathcal{H}_1,\widehat{\mathcal{H}}_0\big\}+\widehat{\pi}_0D_{\text{tr}}P_{\text{tr}}\leq\overline{P}_{\text{av}}, \tag{43}$$

where $\mathbb{E}\{P|\mathcal{H}_0,\widehat{\mathcal{H}}_0\}=\int_0^{\infty}P(y)f_{\widehat{\nu}^*}^0(y)dy$, and the third term in (43) accounts for transmit power used for training symbols. We note that spectrum sensing error affects ATPC through variables β_0, β_1 and $\widehat{\pi}_0$. Also, channel estimation error affects ATPC through variable P.

Now that we have characterized a lower bound on the achievable rates R_{LB} in (37), AIC in (41), and ATPC in (43), we summarize how the four error types, namely, spectrum sensing error, beam detection error, channel estimation error, and beam selection error, affect these expressions. First, spectrum sensing error affects AIC via β_1, both ATPC and R_{LB} via β_0 and β_1. Recall β_0,β_1 depend on $\pi_0,\overline{P}_{\text{fa}},\overline{P}_{\text{d}}$ (see (11)). Second, beam detection error affects AIC via $\overline{\Delta}_{m_{\text{PU}}^*,i}$ and does not have a direct impact on ATPC and R_{LB}. Third, channel estimation error affects both AIC and ATPC via T_{tr}, and R_{LB} via $\widetilde{\alpha}_m^{\ell}$. Fourth, beam selection error impacts AIC, ATPC and R_{LB} via P (which depends on the estimation channel gain of the selected beam).

Having the mathematical expressions for R_{LB}, AIC, ATPC, our goal is to allocate transmission resources such that R_{LB} is maximized, subject to the aforementioned constraints. To determine our optimization variables, we need to examine closely the underlying trade-offs between decreasing average interference and average transmit powers, decreasing four types of errors (i.e., spectrum sensing error, beam detection error, channel estimation error, and beam selection error), and increasing R_{LB}. Within a frame with fixed duration of T_{f} seconds, time is divided between three phases with variable durations: spatial spectrum sensing with duration T_{se}, channel training with duration T_{tr}, and data transmission with duration of T_{d}. Suppose T_{se} increases. On the positive side, spectrum sensing error, beam detection error, and average interference imposed on PU decrease (i.e., for ideal spectrum sensing $\beta_1=0$ in (11) and data transmission from $\mathrm{SU_{tx}}$ to $\mathrm{SU_{rx}}$ does not cause interference on PU). On the negative side, $T_{\mathrm{tr}}+T_{\mathrm{d}}$ decreases, that can lead to increasing channel estimation error (due to decrease in T_{tr}) and/or decreasing R_{LB} (due to decrease in T_{d}). Given T_{se}, as T_{tr} increases, channel estimation error in (22) decreases. However, average interference imposed on PU during transmission of training symbols increases and R_{LB} decreases [4]. Finally, increasing data symbol transmission power P increases R_{LB}, however, it increases average interference and average transmit power. Based on all these existing trade-offs, we seek the optimal $T_{\mathrm{se}}, T_{\mathrm{tr}}, P$ such that R_{LB} in (37) is maximized, subject to AIC and ATPC given in (41) and (43), respectively. In other words, we are interested in solving the following constrained optimization problem

$$
\text{(P1)} \qquad \underset{T_{\mathrm{se}},T_{\mathrm{tr}},P}{\text{Maximize}} \quad R_{\mathrm{LB}}
$$
$$
\text{s.t.:} \quad 0<T_{\mathrm{se}}<T_{\mathrm{f}}-T_{\mathrm{tr}}
$$
$$
T_{\mathrm{tr}}>0, \quad P\geq 0
$$
$$
\text{(41) and (43) are satisfied.}
$$

Before delving into the solution of (P1), we have a remark on how our adopted fading model in Section II-B affects our derivations in this section.

Remark: Our theoretical framework can be extended to the more general Nakagami fading model, however, certain expressions need to be re-derived. In particular, $\overline{P}_{\mathrm{d}}=\mathbb{E}_{\boldsymbol{\psi}}\{P_{\mathrm{d}}\}$ in (10) changes, since the pdf of $\boldsymbol{\psi}$ changes. Also, the conditional pdf of $\widehat{\nu}_m$ given $\{\mathcal{H}_0,\mathcal{H}_\ell\}$ in (23), and the CDF and pdf of $\widehat{\nu}^*$ in (25), (26) change. Consequently, the expressions for Ψ_i^ℓ in (27), $\mathbb{E}\{P|\mathcal{H}_1,\widehat{\mathcal{H}}_0\}$ in (42c), and R_{LB} in (37) must be re-calculated.

VI. Constrained Maximization of Rate Lower Bound

In this section, we address the optimization problem (P1). Taking the second derivative of R_{LB} with respect to (w.r.t.) the optimization variables, we note that (P1) is not jointly concave over T_{se}, T_{tr}, P. However, given T_{se} and T_{tr}, (P1) is concave [5] w.r.t. P. We propose an iterative method based on the block coordinate descent (BCD) algorithm to solve (P1). The underlying principle of the BCD algorithm is that, at each iteration one variable is optimized, while the remaining variables are fixed. The iteration continues until it converges to a stationary point of (P1) [46]. To apply the principle of the BCD algorithm to (P1), we consider the following three steps.

[4]Note that as channel estimation error in (22) decreases, the lower bounds in (36) increase. However, this logarithmic increase is dominated by the linear decrease of D_{d} in (37), which leads into a decrease in R_{LB}.

[5]The cost function of (P1) given in (37) depends on P through the two logarithms, that can be viewed, in terms of P, as $(1+\frac{aP}{bP+c})$, where a,b,c are positive. Since the arguments of these logarithms are concave, R_{LB} is also concave w.r.t. P.

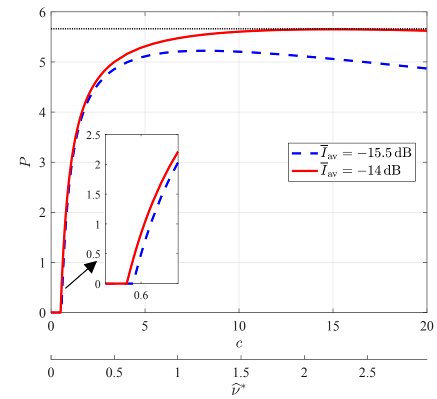


Fig. 4: The optimized P obtained from (45a) versus $\widehat{\nu}^*$ (and c) for $\overline{P}_{\mathrm{av}}=2$ dB.

- **Step (i)**: given T_{se}, T_{tr}, we optimize P using the Lagrangian method. The Lagrangian is

$$
\mathcal{L}=-R_{\mathrm{LB}}+\mu\big[\text{LHS of (41)}-\overline{I}_{\mathrm{av}}\big]+\lambda\big[\text{LHS of (43)}-\overline{P}_{\mathrm{av}}\big], \tag{44}
$$

in which LHS stands for left-hand side, λ and μ are the nonnegative Lagrange multipliers, associated with the ATPC and AIC, respectively. Therefore, the optimal P that minimizes (44) is the solution to the Karush-Kuhn-Tucker (KKT) optimality necessary and sufficient conditions. The KKT conditions are the first derivatives of $\mathcal{L}$ w.r.t. P,μ,λ being equal to zero, i.e., $\partial\mathcal{L}/\partial P=0, \partial\mathcal{L}/\partial\mu=0, \partial\mathcal{L}/\partial\lambda=0$. We have

$$
-\frac{1}{\ln(2)}\sum_{\ell=0}^{1}\beta_\ell\sum_{i=1}^{M}\frac{y\left(\sigma_{\mathrm{q}}^2+\ell\sigma_{\mathrm{p}}^2\right)f_{\widehat{\nu}_i}^{\ell}(y)}{\sigma_{\eta_{i,\ell}}^2\left(yP+\sigma_{\eta_{i,\ell}}^2\right)}\prod_{\substack{m=1\\m\neq i}}^{M}F_{\widehat{\nu}_m}^{\ell}(y)
+\lambda\Big[\beta_0 f_{\widehat{\nu}^*}^{0}(y)+\beta_1 f_{\widehat{\nu}^*}^{1}(y)\Big]+\mu b_0 f_{\widehat{\nu}^*}^{1}(y)=0, \tag{45a}
$$

$$
\mu\Big|\text{LHS of (41)}-\overline{I}_{\mathrm{av}}\Big|=0, \tag{45b}
$$

$$
\lambda\Big|\text{LHS of (43)}-\overline{P}_{\mathrm{av}}\Big|=0. \tag{45c}
$$

The closed-form analytical solution for (45) cannot be found. Hence, we solve these equations numerically for every realization of $\widehat{\nu}^*$, via the following iterative method. We first initialize the Lagrangian multipliers μ and λ and then find P using (45a), and verify that it satisfies (45b), (45c). Next, we update μ and λ using the subgradient method. Using the updated μ and λ, we find P again using (45a). We repeat this procedure until μ and λ converge (i.e., a pre-determined stopping criterion is met).

- **Step (ii)**: given P and T_{tr}, we optimize T_{se}. The optimal T_{se} is the solution of the equation $\partial R_{\mathrm{LB}}/\partial T_{\mathrm{se}}=0$. In Appendix A, we show that this equation has one solution in the interval $(0,T_{\mathrm{f}}-T_{\mathrm{tr}})$. This solution can be found using numerical search methods (e.g., bisection method).
- **Step (iii)**: given P and T_{se}, we optimize T_{tr}, via solving $\partial R_{\mathrm{LB}}/\partial T_{\mathrm{tr}}=0$. In Appendix B, we show that this equation has one solution in the interval $(0,T_{\mathrm{f}}-T_{\mathrm{se}})$, which can be found numerically using search methods (e.g., bisection method).

TABLE II: $\Pr(\widehat{\nu}^* \geq c\, m_{\widehat{\nu}^*})$ in terms of c, given $m_{\widehat{\nu}^*} = 0.1484$.

c	$\Pr(\widehat{\nu}^* \geq c\, m_{\widehat{\nu}^*})$
4	3.01×10^{-3}
8	7.04×10^{-6}
12	1.54×10^{-8}
16	4.87×10^{-11}

To gain an insight on the solution of (P1), we look into the behavior of the optimized P versus the realizations of the estimated channel gain $\widehat{\nu}^*$. Fig. 4 illustrates the optimized P versus $\widehat{\nu}^*$ (and c, where $\widehat{\nu}^* = c\, m_{\widehat{\nu}^*}$ and $m_{\widehat{\nu}^*}$ is the mean of $\widehat{\nu}^*$) for $\overline{I}_{\rm av} = -15.5, -14$ dB and other simulation parameters given in Table III. For these parameters $m_{\widehat{\nu}^*} = 0.1484$. We observe that the optimized P for very small $\widehat{\nu}^*$ (when $\widehat{\nu}^*$ is smaller than a cut-off threshold $\zeta = 3.5\, m_{\widehat{\nu}^*}$) is zero. As $\widehat{\nu}^*$ increases the optimized P increases gradually until it reaches a maximum value. As $\widehat{\nu}^*$ increases further, the optimized P decreases, until it reaches a minimum value for very large $\widehat{\nu}^*$ (when $\widehat{\nu}^* > 85\, m_{\widehat{\nu}^*}$), not shown in the figure. Comparing the curves for $\overline{I}_{\rm av} = -15.5$ dB and $\overline{I}_{\rm av} = -14$ dB, we note that the optimized P decays faster (after it reaches its maximum value) for lower $\overline{I}_{\rm av}$. Moreover, the cut-off threshold ζ is lower for higher $\overline{I}_{\rm av}$. The behavior of the optimized P versus $\widehat{\nu}^*$ is different from our intuitive expectation that expects to see the optimized P increases monotonically as $\widehat{\nu}^*$ increases. We explore this by examining the optimized P, which satisfies (45a).

Although for general M the optimized P does not have a closed form expression, for $M = 1$ and under a simplifying assumption[6] it can approximated as follows:

$$P \approx \left[\frac{F + \sqrt{\Upsilon}}{2}\right]^+, \tag{46}$$

$$F = \frac{\beta_0 W(\widehat{\nu}^*) + \beta_1}{\ln(2)\Big[\lambda\big(\beta_0 W(\widehat{\nu}^*) + \beta_1\big) + \mu b_0\Big]} - \frac{2\sigma_{\rm q}^2 + \sigma_{\rm p}^2}{\widehat{\nu}^*},$$

$$\Upsilon = F^2 - \frac{4}{\widehat{\nu}^*}\left(\frac{\sigma_{\rm q}^2(\sigma_{\rm q}^2 + \sigma_{\rm p}^2)}{\widehat{\nu}^*} - \frac{\big(\beta_0 W(\widehat{\nu}^*) + \beta_1\big)\sigma_{\rm q}^2 + \beta_1 \sigma_{\rm p}^2}{\ln(2)\Big[\lambda\big(\beta_0 W(\widehat{\nu}^*) + \beta_1\big) + \mu b_0\Big]}\right).$$

where $W(\widehat{\nu}^*) = f^0_{\widehat{\nu}^*}(\widehat{\nu}^*)/f^1_{\widehat{\nu}^*}(\widehat{\nu}^*) = \widehat{\alpha}^1/\widehat{\alpha}^0\ e^{-\widehat{\nu}^*(\frac{1}{\widehat{\alpha}^0} - \frac{1}{\widehat{\alpha}^1})}$. Considering (21) we realize that $\widehat{\alpha}^0 < \widehat{\alpha}^1$. This implies as $\widehat{\nu}^*$ increases, $W(\widehat{\nu}^*)$ and Υ decrease. However, the behavior of F changes, i.e., F increases until it reaches a maximum value. As $\widehat{\nu}^*$ increases further, F decreases. Considering (46) we note that the behavior of P (in terms of $\widehat{\nu}^*$) is dominated by the behavior of F. In the ideal scenario when there is no channel estimation error, we have $\widehat{\alpha}^0 = \widehat{\alpha}^1 = \alpha$ and $W(\widehat{\nu}^*) = 1$, F monotonically increases and Υ deceases, i.e., P in (46) monotonically increases as $\widehat{\nu}^*$ increases, which is what we intuitively expect.

The optimized P we discussed so far requires solving (45) several times for each realization of $\widehat{\nu}^*$. Integrating the insights we have gained into how this optimized P varies in terms of $\widehat{\nu}^*$, we propose two transmit power control schemes that are simpler to implement and yield achievable rate lower bounds that are very close to the maximized $R_{\rm LB}$ values in (P1). Since $\Pr(\widehat{\nu}^* \geq c\, m_{\widehat{\nu}^*})$ is very small for $c \geq 8$ (see Table II), we focus on the regime when $\widehat{\nu}^* < 8\, m_{\widehat{\nu}^*}$ and develop two schemes, dubbed here Scheme 1 and Scheme 2, that mimic the behavior of the optimized P in this regime.

[6]We assume that the optimized $T_{\rm tr}$ is large enough such that $\widetilde{\alpha}P + \sigma_{\rm q}^2 \approx \sigma_{\rm q}^2$. This assumption allows us to approximate (45a) for $M = 1$ as a quadratic polynomial in P (originally a polynomial of degree 4 in P) and find a closed-form expression for P.

A. Scheme 1

For Scheme 1, when the spectrum is sensed idle, $\text{SU}_{\rm tx}$ sends data to $\text{SU}_{\rm rx}$ over the selected sector $i = m^*_{\rm SR}$ according to the following rule:

$$P_{\rm S_1} = \begin{cases} \Pi_1, & \text{if } \ \widehat{\nu}^* \geq \zeta_1 \\ 0, & \text{if } \ \widehat{\nu}^* < \zeta_1 \end{cases} \tag{47}$$

i.e., when $\widehat{\nu}^*$ is less than a cut-off threshold ζ_1, $\text{SU}_{\rm tx}$ remains silent, when $\widehat{\nu}^*$ is larger than ζ_1, $\text{SU}_{\rm tx}$ lets its transmit power be equal to constant Π_1. The parameter Π_1 can be found in terms of $T_{\rm se}, T_{\rm tr}, \zeta_1$, via enforcing AIC in (41) and ATPC in (43) as the following:

$$\Pi_1 = \frac{1}{D_{\rm d}} \min\left\{\frac{\overline{P}_{\rm av} - \widehat{\pi}_0 D_{\rm tr} P_{\rm tr}}{\sum_{\ell=0}^{1} \beta_\ell\big(1 - F^\ell_{\widehat{\nu}^*}(\zeta_1)\big)}, \frac{\overline{I}_{\rm av} - u_0 D_{\rm tr} P_{\rm tr}}{b_0\big(1 - F^1_{\widehat{\nu}^*}(\zeta_1)\big)}\right\}. \tag{48}$$

Let $R_{\rm S_1}$ denote the lower bound on the achievable rates when $\text{SU}_{\rm tx}$ adopts the power control scheme in (47). We find $R_{\rm S_1}$ expression by substituting $P_{\rm S_1}$ in (37) and taking expectation w.r.t. $\widehat{\nu}^*$. This expression is given in (49) where $\text{SNR}^0_i = \frac{\Pi_1}{\widehat{\alpha}^0_i + \sigma_{\rm q}^2}$, $\text{SNR}^1_i = \frac{\Pi_1}{\widehat{\alpha}^1_i + \sigma_{\rm q}^2 + \sigma_{\rm p}^2}$ and $\text{Ei}(\cdot)$ is the exponential integral. With this transmit power scheme, we consider a modified problem to (P1), where the lower bound $R_{\rm S_1}$ in (49) is maximized (subject to the same constraints) and the optimization variables are $T_{\rm se}, T_{\rm tr}, \zeta_1$. To solve this modified problem, we use an iterative method based on the BCD algorithm and implement the following three steps: **Step (i)**, given $T_{\rm se}, T_{\rm tr}$, we optimize ζ_1, via maximizing $R_{\rm S_1}$, using bisection search method. **Step (ii)**, given $\zeta_1, T_{\rm tr}$, we optimize $T_{\rm se}$, using bisection search method. **Step (iii)**, given $\zeta_1, T_{\rm se}$, we optimize $T_{\rm tr}$, using bisection search method. In Section VII we numerically compare the maximized $R_{\rm LB}$ in (P1) and the maximized $R_{\rm S_1}$.

B. Scheme 2

For Scheme 2, when the spectrum is sensed idle, $\text{SU}_{\rm tx}$ sends data symbols to $\text{SU}_{\rm rx}$ over the selected sector $i = m^*_{\rm SR}$ according to the following rule:

$$P_{\rm S_2} = \begin{cases} \Pi_2\big(1 - \frac{\zeta_2}{\widehat{\nu}^*}\big), & \text{if } \ \widehat{\nu}^* \geq \zeta_2 \\ 0, & \text{if } \ \widehat{\nu}^* < \zeta_2 \end{cases} \tag{50}$$

Different from Scheme 1, in the Scheme 2 when $\widehat{\nu}^*$ exceeds the cut-off threshold ζ_2, $\text{SU}_{\rm tx}$ transmits at a variable power. The power level increases as $\widehat{\nu}^*$ increases, until it reaches its maximum value of Π_2, i.e., $\lim_{\widehat{\nu}^* \to \infty} P_{\rm S_2} = \Pi_2$. The

$$R_{\mathrm{S}_1} = \frac{D_{\mathrm{d}}}{\ln(2)} \sum_{\ell=0}^{1} \beta_\ell \sum_{j=1}^{M} \Big[Y\big(\widehat{\alpha}_j^\ell, \mathrm{SNR}_j^\ell\big) + \sum_{\substack{m=1\\ m\neq j}}^{M} (-1)^m \sum\nolimits_m Y\big(d_{j,m}^\ell, \mathrm{SNR}_j^\ell\big) \Big] \tag{49}$$

$$Y(a,b) = \int_{\zeta_1}^{\infty} \ln(1+bx) \frac{1}{a} e^{\frac{-x}{a}} dx = e^{-\zeta_1/a} \ln\big(1+b\zeta_1\big) - e^{1/ab}\, \mathrm{Ei}\big(-\zeta_1/a - 1/ab\big), \qquad d_{j,m}^\ell = \left(A_{k_1:k_m}^\ell + \frac{1}{\widetilde{\alpha}_j^\ell} \right)^{-1}.$$

parameter Π_2 can be found in terms of $T_{\mathrm{se}}, T_{\mathrm{tr}}, \zeta_2$, via enforcing AIC in (41) and ATPC in (43) as the following:

$$\Pi_2 = \frac{1}{D_{\mathrm{d}}} \min\left\{ \frac{\overline{P}_{\mathrm{av}} - \widehat{\pi}_0 D_{\mathrm{tr}} P_{\mathrm{tr}}}{\sum_{\ell=0}^{1} \beta_\ell \big[1 - G^\ell(\zeta_2)\big]}, \frac{\overline{I}_{\mathrm{av}} - u_0 D_{\mathrm{tr}} P_{\mathrm{tr}}}{b_0 \big[1 - G^1(\zeta_2)\big]} \right\}, \tag{51}$$

where $G^\ell(\zeta_2) = F_{\widehat{\nu}^*}^\ell(\zeta_2) + \zeta_2 T^\ell(\zeta_2)$ and

$$T^\ell(\zeta_2) = \mathbb{E}\left\{ \frac{1}{\widehat{\nu}^*} \,\Big|\, \widehat{\nu}^* \geq \zeta_2, \mathcal{H}_\ell \right\} = \int_{\zeta_2}^{\infty} \frac{f_{\widehat{\nu}^*}^\ell(y)}{y} dy$$
$$= \sum_{m=1}^{M} (-1)^m \sum\nolimits_m A_{j_1:j_m}^\ell \mathrm{Ei}\big(-\zeta_2 A_{j_1:j_m}^\ell\big). \tag{52}$$

Let R_{S_2} represent the lower bound on the achievable rates when $\mathrm{SU}_{\mathrm{tx}}$ adopts the power control scheme in (50). We find R_{S_2} by substituting P_{S_2} in (37) and taking expectation w.r.t. $\widehat{\nu}^*$. With this transmit power scheme, we consider a modified problem to (P1), where the lower bound R_{S_2} is maximized (subject to the same constraints) and the optimization variables are $T_{\mathrm{se}}, T_{\mathrm{tr}}, \zeta_2$. To solve this modified problem, we use an iterative method based on the BCD algorithm and implement the following three steps: **Step (i)**, given $T_{\mathrm{tr}}, T_{\mathrm{se}}$, we optimize ζ_2, via maximizing R_{S_2}, using bisection search method. **Step (ii)**, given ζ_2, T_{tr}, we optimize T_{se}, using bisection search method. **Step (iii)**, given ζ_2, T_{se}, we optimize T_{tr}, using bisection search method. In Section VII we numerically compare the maximized R_{LB} in (P1) and the maximized R_{S_2}. Note that the closed-form expression for R_{S_2} cannot be obtained.

C. Discussion on Computational Complexity of Proposed Algorithms

In the following, we discuss the computational complexity of the three proposed algorithms, namely, the first algorithm in Section VI, Scheme 1 in Section VI-A, and Scheme 2 in Section VI-B.

The first algorithm consists of three steps. We discuss the computational complexity of each step. **Step (i)**: given T_{se}, T_{tr}, we find P via solving (45a), (45b), (45c) numerically. In particular, noting that y in (45a) is positive, we partition the real positive line into N_y intervals. Given y is in one of these N_y intervals, we initialize the Lagrangian multipliers μ and λ and then solve (45a) for P using bisection method. The computational complexity of bisection method to provide an ϵ_p-accurate solution for each of these N_y intervals is $\mathcal{O}(\log(1/\epsilon_p))$ [47], [48]. Hence, the computational complexity for solving (45a) N_y times is $\mathcal{O}(N_y \log(1/\epsilon_p))$. Moving on to (45b) and (45c), we need to compute LHS of (41) and (43), respectively, which requires calculating the conditional expectations $\mathbb{E}\{P|\mathcal{H}_1, \widehat{\mathcal{H}}_0\}$ and $\mathbb{E}\{P|\mathcal{H}_0, \widehat{\mathcal{H}}_0\}$ and integrating over y. Hence, the computational complexity for computing (45b) and (45c) is $\mathcal{O}(N_y)$. Since $N_y \ll N_y \log(1/\epsilon_p)$, we can neglect the computational complexity of solving (45b), (45c), with respect to that of solving (45a). Hence, the computational complexity of solving (45), given μ and λ, is $\mathcal{O}(N_y \log(1/\epsilon_p))$. Next, we update μ and λ using the subgradient method. Using the updated μ and λ, we solve (45a) for P again. We repeat this procedure until both μ and λ converge. The computational complexity to get $\epsilon_{\mathcal{L}}$-convergence for μ and λ is $\mathcal{O}(S_1)$, where $S_1 = (N_y \log(1/\epsilon_p))/\epsilon_{\mathcal{L}}$. **Step (ii)**: given P and T_{tr}, we find T_{se} using bisection search method. The computational complexity of bisection search method to provide an ϵ_{se}-accurate solution is $\mathcal{O}(S_2)$, where $S_2 = \log(1/\epsilon_{\mathrm{se}})$. **Step (iii)**: given P and T_{se}, we find T_{tr} using bisection search method. The computational complexity of bisection search method to provide an ϵ_{tr}-accurate solution is $\mathcal{O}(S_3)$, where $S_3 = \log(1/\epsilon_{\mathrm{tr}})$. At each iteration of **Step (iii)**, we execute **Step (ii)** and at each iteration of **Step (ii)**, we execute **Step (i)**. Hence, the overall computational complexity of the first algorithm is $\mathcal{O}(S_1 S_2 S_3)$.

Scheme 1: Similar to the first algorithm, Scheme 1 consists of three steps. At the first step, given T_{se}, T_{tr}, we optimize ζ_1 using bisection search method. The computational complexity of bisection search method to provide an ϵ_ζ-accurate solution is $\mathcal{O}(\log(1/\epsilon_\zeta))$. The second and third steps are exactly the same as **Step (ii)** and **Step (iii)** in the first algorithm. Hence, the overall computational complexity of Scheme 1 is $\mathcal{O}(S_2 S_3 \log(1/\epsilon_\zeta))$.

Scheme 2: Similar to the first algorithm, Scheme 2 consists of three steps. At the first step, given T_{se}, T_{tr}, we optimize ζ_2 using bisection search method. The computational complexity of bisection search method to provide an ϵ_ζ-accurate solution is $\mathcal{O}(\log(1/\epsilon_\zeta))$. The computational complexity of integrating over y in (37) within each iteration of bisection search method is $\mathcal{O}(N_y)$. Hence, the computational complexity for the first step of Scheme 2 is $\mathcal{O}(N_y \log(1/\epsilon_\zeta))$. The second and third steps are exactly the same as **Step (ii)** and **Step (iii)** in the first algorithm. Hence, the overall computational complexity of Scheme 2 is $\mathcal{O}(N_y S_2 S_3 \log(1/\epsilon_\zeta))$.

Comparing the computational complexity of these three schemes, it is clear that Scheme 2 has a higher computational complexity than that of Scheme 1. Under the assumption $\epsilon_\zeta = \epsilon_{\mathcal{L}} = \epsilon_p$, we find that the first scheme has the highest and Scheme 1 has the lowest computational complexity.

VII. Simulation Results

We corroborate our analysis on constrained maximization of achievable rate lower bounds with Matlab simulations. Our simulation parameters are given in Table III. We start by illustrating the the behavior of our proposed power allocation schemes versus $\widehat{\nu}^*$. Fig. 5 shows the optimized P obtained

TABLE III: Simulation Parameters

Parameter	Value	Parameter	Value	Parameter	Value
A_0	0.98	γ_{ss}	0.1	$\sigma_{\text{w}}^2, \sigma_{\text{q}}^2$	0.5
A_1	0.02	$\gamma, \gamma_{\text{sp}}$	0.5	P_{p}	0.5 watts
$\phi_{3\text{dB}}$	20°	π_0	0.7	T_{f}	30 ms
ϕ_1	-55°	$\overline{P}_{\text{d}}$	0.85	P_{tr}	2 watts
ϕ_2	$+55^{\circ}$	M	7		

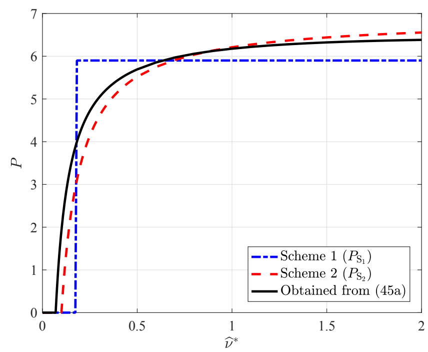
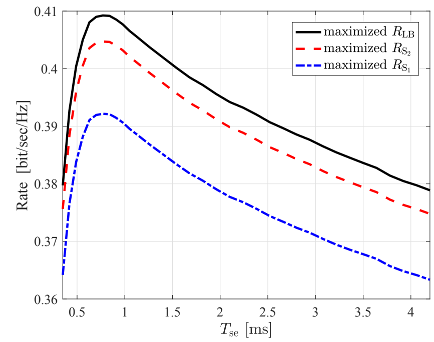
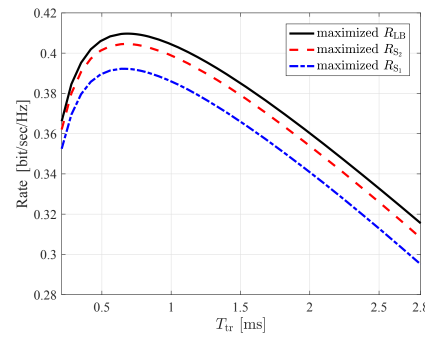


Fig. 5: P versus $\widehat{\nu}^*$ for $\overline{P}_{\text{av}} = 2\,\text{dB}, \overline{I}_{\text{av}} = -12\,\text{dB}$.



(a)



(b)

Fig. 6: (a) Rate versus T_{se}, (b) Rate versus T_{tr} for $\overline{P}_{\text{av}} = 2\,\text{dB}, \overline{I}_{\text{av}} = -15\,\text{dB}$.

by solving (45a) and the two proposed suboptimal schemes P_{S_2} and P_{S_1} versus $\widehat{\nu}^*$. We observe that P_{S_2} and P_{S_1} mimic the behavior of the optimized P. Furthermore, for the cut-off thresholds we have $\zeta < \zeta_1 < \zeta_2$.

Next, we explore the effect of spatial spectrum sensing duration T_{se} on the achievable rate lower bounds of our system. Fig. 6a shows the maximized $R_{\text{LB}}, R_{\text{S}_2}$ and R_{S_1} (which we refer in the figures to as "Rate") versus T_{se}. To plot this figure, we maximize the bounds w.r.t. only T_{tr} and P, subject to ATPC and AIC. We note that for all T_{se} values we have $R_{\text{LB}} > R_{\text{S}_2} > R_{\text{S}_1}$. We observe that the achievable rates always have a maximum in the interval $(0, T_{\text{f}} - T_{\text{tr}})$. For the simulation parameters in Table III the optimized $T_{\text{se}} = 0.75\,\text{ms} = 2.5\%\,T_{\text{f}}$. Also, Scheme 2 yields a higher achievable rate than that of Scheme 1, because its corresponding power P_{S_2} fits better to the optimized power P obtained from solving (45a). The achievable rate R_{S_2} is very close to R_{LB} and we do not have a significant performance loss if we choose the simple transmit power control scheme in (50).

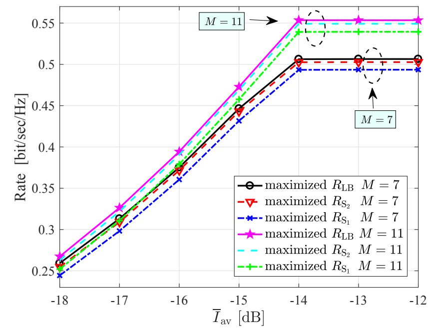
To investigate the effect of channel training duration T_{tr} on the achievable rate lower bounds, we plot Fig. 6b which illustrates the maximized $R_{\text{LB}}, R_{\text{S}_2}$ and R_{S_1} versus T_{tr}. To plot this figure, we maximize the bounds w.r.t. only T_{se} and P, subject to ATPC and AIC. For all T_{tr} values we have $R_{\text{LB}} > R_{\text{S}_2} > R_{\text{S}_1}$. We observe that the achievable rates always have a maximum in the interval $(0, T_{\text{f}} - T_{\text{se}})$. For the simulation parameters in Table III the optimized $T_{\text{tr}} = 0.67\,\text{ms} = 2.23\%\,T_{\text{f}}$. Comparing Fig. 6b and Fig. 6a, we notice that the achievable rates are more sensitive to the variations of T_{tr} compared to that of T_{se}. To be more specific, considering Fig. 6a and Fig. 6b, suppose we choose T_{se} and T_{tr} values that are different from their corresponding maximum values by 20%, i.e., $\Delta T_{\text{se}} = 20\%, \Delta T_{\text{tr}} = 20\%$. Then

$$\Delta R_{\text{LB}}/\Delta T_{\text{tr}} > \Delta R_{\text{LB}}/\Delta T_{\text{se}},$$
$$\Delta R_{\text{S}_2}/\Delta T_{\text{tr}} > \Delta R_{\text{S}_2}/\Delta T_{\text{se}},$$
$$\Delta R_{\text{S}_1}/\Delta T_{\text{tr}} > \Delta R_{\text{S}_1}/\Delta T_{\text{se}}.$$

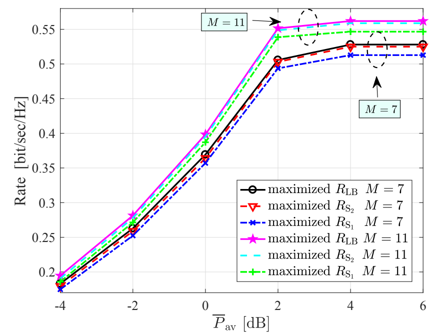
These indicate that proper allocation of T_{tr} is more important than that of T_{se}, for providing higher achievable rates in our system.

To explore the effects of the number of beams M and $\overline{I}_{\text{av}}$ on the achievable rate lower bounds, Fig. 7a illustrates the maximized $R_{\text{LB}}, R_{\text{S}_2}, R_{\text{S}_1}$ versus $\overline{I}_{\text{av}}$ for $M = 7, 11$ and $\overline{P}_{\text{av}} = 2\,\text{dB}$. We observe that as M increases a higher rate can be achieved. For all M and $\overline{I}_{\text{av}}$ values we have $R_{\text{LB}} > R_{\text{S}_2} > R_{\text{S}_1}$. We realize that as $\overline{I}_{\text{av}}$ increases from $-18\,\text{dB}$ to $-14\,\text{dB}$, the achievable rates are monotonically increasing and the AIC is dominant. However, as $\overline{I}_{\text{av}}$ increases beyond $-14\,\text{dB}$, the achievable rates remain unchanged and the ATPC is dominant. Fig. 7b illustrates the maximized $R_{\text{LB}}, R_{\text{S}_2}, R_{\text{S}_1}$ versus $\overline{P}_{\text{av}}$ for $M = 7, 11$ and $\overline{I}_{\text{av}} = -14\,\text{dB}$. The behaviors of the achievable rates in terms of M are the same as Fig. 7a. We note that as $\overline{P}_{\text{av}}$ increases from $-4\,\text{dB}$ to $2\,\text{dB}$, the achievable rates are monotonically increasing and the ATPC is dominant. However, as $\overline{P}_{\text{av}}$ increases beyond $2\,\text{dB}$, the achievable rates remain unchanged and the AIC is dominant.

We also consider outage probability as another performance metric to evaluate our system. We define the outage probability as the probability of SU_{tx} not transmitting data symbols due to the weak SU_{tx}-SU_{rx} channel when the spectrum is sensed idle, i.e., $P_{\text{out}} = \Pr\left\{P = 0\,|\,\widehat{\mathcal{H}}_0\right\}$. This probability can be



(a)



(b)

Fig. 7: (a) Rate versus $\overline{I}_{\text{av}}$ for $M = 7, 11$ and $\overline{P}_{\text{av}} = 2\,\text{dB}$, (b) Rate versus $\overline{P}_{\text{av}}$ for $M = 7, 11$ and $\overline{I}_{\text{av}} = -14\,\text{dB}$.

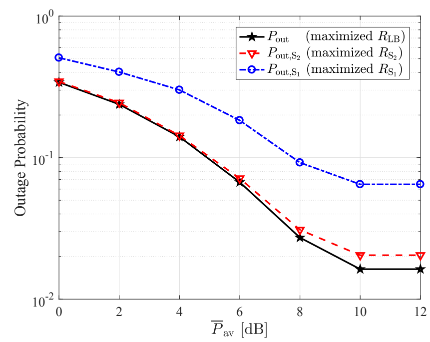


Fig. 8: P_{out} versus $\overline{P}_{\text{av}}$ for $\overline{I}_{\text{av}} = -8\,\text{dB}$.

directly obtained using the CDF of $\widehat{\nu}^*$ evaluated at the cut-off threshold as the following

$$\begin{aligned} P_{\text{out}} &= \Pr(\widehat{\nu}^* \le \zeta | \widehat{\mathcal{H}}_0) = \omega_0 F^0_{\widehat{\nu}^*}(\zeta) \; + \omega_1 F^1_{\widehat{\nu}^*}(\zeta), \\ P_{\text{out,S}_1} &= \Pr(\widehat{\nu}^* \le \zeta_1 | \widehat{\mathcal{H}}_0) = \omega_0 F^0_{\widehat{\nu}^*}(\zeta_1) + \omega_1 F^1_{\widehat{\nu}^*}(\zeta_1), \\ P_{\text{out,S}_2} &= \Pr(\widehat{\nu}^* \le \zeta_2 | \widehat{\mathcal{H}}_0) = \omega_0 F^0_{\widehat{\nu}^*}(\zeta_2) + \omega_1 F^1_{\widehat{\nu}^*}(\zeta_2). \end{aligned}$$

Fig. 8 illustrates $P_{\text{out}}, P_{\text{out,S}_2}, P_{\text{out,S}_1}$ versus $\overline{P}_{\text{av}}$ for $\overline{I}_{\text{av}} = -8\,\text{dB}$. We observe that as $\overline{P}_{\text{av}}$ increases the outage probabilities decrease. Moreover, for a given $\overline{P}_{\text{av}}$ we have $P_{\text{out}} < P_{\text{out,S}_2} < P_{\text{out,S}_1}$. This is consistent with Fig. 5 which shows for a given $\overline{P}_{\text{av}}$, we have $\zeta < \zeta_1 < \zeta_2$. Combined this with the fact that the CDF $F_{\widehat{\nu}^*}(\cdot)$ is an increasing function of its argument, we reach the conclusion that $P_{\text{out}} < P_{\text{out,S}_2} < P_{\text{out,S}_1}$.

VIII. Conclusions

We considered an opportunistic CR system consisting of a PU, SU_{tx}, and SU_{rx}, where SU_{tx} is equipped with a RA that has M beams, and there is an error-free low-rate feedback channel from SU_{rx} to SU_{tx}. We proposed a system design for integrated sector-based spatial spectrum sensing and sector-based data symbol communication. We studied the entangled effects of spectrum sensing error, channel estimation error, and beam detection and beam selection errors (introduced by the RA), on the system achievable rates. We formulated a constrained optimization problem, where a lower bound on the achievable rate of SU_{tx}–SU_{rx} link is maximized, subject to ATPC and AIC, with the optimization variables being the durations of spatial spectrum sensing T_{se} and channel training T_{tr} as well as data symbol transmission power at SU_{tx}. Moreover, we proposed two alternative power adaptation schemes that are simpler to implement. We solved the proposed constrained optimization problems using iterative methods based on the BCD algorithm. Our simulation results demonstrate that one can increase the achievable rates of SU_{tx}–SU_{rx} link significantly, via implementing these optimizations, while maintaining the ATPC and AIC. They also showed that the achievable rates obtained from employing simple Schemes 1 and 2 are very close to the one produced by the optimized transmit power. Our numerical results also showed that between optimizing T_{se} and T_{tr}, optimizing the latter has a larger effect on increasing the achievable rates in our system.

Appendix A
Showing that $\partial R_{\text{LB}}/\partial T_{\text{se}} = 0$ has one solution in the interval $(0, T_{\text{f}} - T_{\text{tr}})$

Let $R_{\text{LB}} = C_0 + C_1$ where $C_0 = D_{\text{d}}\beta_0 R_0$ and $C_1 = D_{\text{d}}\beta_1 R_1$. To calculate $\partial R_{\text{LB}}/\partial T_{\text{se}}$ we need the following derivatives:

$$\begin{aligned} \frac{\partial C_0}{\partial T_{\text{se}}} &= R_0 \left[\beta_0 \frac{\partial D_{\text{d}}}{\partial T_{\text{se}}} + D_{\text{d}} \frac{\partial \beta_0}{\partial T_{\text{se}}}\right] = R_0 \left[\frac{-\beta_0}{T_{\text{f}}} + D_{\text{d}} \frac{\partial \beta_0}{\partial T_{\text{se}}}\right], \\ \frac{\partial C_1}{\partial T_{\text{se}}} &= R_1 \left[\beta_1 \frac{\partial D_{\text{d}}}{\partial T_{\text{se}}} + D_{\text{d}} \frac{\partial \beta_1}{\partial T_{\text{se}}}\right] = R_1 \left[\frac{-\beta_1}{T_{\text{f}}} + D_{\text{d}} \frac{\partial \beta_1}{\partial T_{\text{se}}}\right]. \end{aligned}$$

Recall $\beta_0 = \pi_0(1 - \overline{P}_{\text{fa}})$ and $\beta_1 = \pi_1(1 - \overline{P}_{\text{d}})$ in (11). We assume $\overline{P}_{\text{d}}$ is given, hence $\partial\beta_1/\partial T_{\text{se}} = 0$. On the other hand, $\overline{P}_{\text{fa}}$ in (8) is variable w.r.t. T_{se}, and hence we have

$$\frac{\partial \beta_0}{\partial T_{\text{se}}} = \pi_0 f_{\text{TW2}}\Big(\frac{\eta - \theta_{\text{sen}}}{\sigma_{\text{sen}}}\Big) \; \frac{\partial}{\partial T_{\text{se}}}\Big(\frac{\eta - \theta_{\text{sen}}}{\sigma_{\text{sen}}}\Big) \tag{53}$$

where f_{TW2} denotes the pdf of the Tracy-Widom distribution of order 2, and, $\theta_{\text{sen}}, \sigma_{\text{sen}}$ are given in (9). Evaluating $\frac{\partial C_0}{\partial T_{\text{se}}}$ and $\frac{\partial C_1}{\partial T_{\text{se}}}$ when $T_{\text{se}} \to 0$ we have

$$\begin{aligned} \lim_{T_{\text{se}}\to 0} \frac{\partial C_0}{\partial T_{\text{se}}} = &\lim_{T_{\text{se}}\to 0} \frac{-\beta_0}{T_{\text{f}}} R_0 \\ &+ \frac{(T_{\text{f}} - T_{\text{tr}})}{T_{\text{f}}} R_0 \underbrace{\Big(\lim_{T_{\text{se}}\to 0} \frac{\partial \beta_0}{\partial T_{\text{se}}}\Big)}_{=+\infty} = +\infty, \end{aligned} \tag{54a}$$

$$\begin{aligned} \lim_{T_{\text{se}}\to 0} \frac{\partial C_1}{\partial T_{\text{se}}} = &\lim_{T_{\text{se}}\to 0} \frac{-\beta_1}{T_{\text{f}}} R_1 \\ &+ \frac{(T_{\text{f}} - T_{\text{tr}})}{T_{\text{f}}} R_1 \underbrace{\Big(\lim_{T_{\text{se}}\to 0} \frac{\partial \beta_1}{\partial T_{\text{se}}}\Big)}_{=0} < 0. \end{aligned} \tag{54b}$$

Evaluating $\frac{\partial C_0}{\partial T_{\rm se}}$ and $\frac{\partial C_1}{\partial T_{\rm se}}$ when $T_{\rm se}\to T_{\rm f}-T_{\rm tr}$ we have

$$\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{\partial C_0}{\partial T_{\rm se}} = \lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{-\beta_0}{T_{\rm f}}R_0 + R_0\underbrace{\Big(\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}} D_{\rm d}\Big)}_{=0}\Big(\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{\partial \beta_0}{\partial T_{\rm se}}\Big) < 0, \quad (55a)$$

$$\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{\partial C_1}{\partial T_{\rm se}} = \lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{-\beta_1}{T_{\rm f}}R_1 + R_1\underbrace{\Big(\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}} D_{\rm d}\Big)}_{=0}\Big(\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{\partial \beta_1}{\partial T_{\rm se}}\Big) < 0. \quad (55b)$$

The inequalities in (54a) and (54b) show that $\lim_{T_{\rm se}\to 0}\frac{\partial R_{\rm LB}}{\partial T_{\rm se}} > 0$. On the other hand, the inequalities in (55a) and (55b) show that $\lim_{T_{\rm se}\to T_{\rm f}-T_{\rm tr}}\frac{\partial R_{\rm LB}}{\partial T_{\rm se}} < 0$. Together, these indicate that the equation $\partial R_{\rm LB}/\partial T_{\rm se}=0$ has one solution in the interval $(0, T_{\rm f}-T_{\rm tr})$. This solution can be found using bisection search method.

Appendix B
Showing that $\partial R_{\rm LB}/\partial T_{\rm tr}=0$ has one solution in the interval $(0, T_{\rm f}-T_{\rm se})$

To calculate $\partial R_{\rm LB}/\partial T_{\rm tr}$ we need the following derivatives:

$$\frac{\partial C_0}{\partial T_{\rm tr}} = \beta_0\Big[D_{\rm d}\frac{\partial R_0}{\partial T_{\rm tr}} + \frac{\partial D_{\rm d}}{\partial T_{\rm tr}}R_0\Big] = \beta_0\Big[D_{\rm d}\sum_{m=1}^{M}\frac{\partial R_0}{\partial \widehat{\alpha}_m^0}\frac{\partial \widehat{\alpha}_m^0}{\partial T_{\rm tr}} - \frac{R_0}{T_{\rm f}}\Big],$$

$$\frac{\partial C_1}{\partial T_{\rm tr}} = \beta_1\Big[D_{\rm d}\frac{\partial R_1}{\partial T_{\rm tr}} + \frac{\partial D_{\rm d}}{\partial T_{\rm tr}}R_1\Big] = \beta_1\Big[D_{\rm d}\sum_{m=1}^{M}\frac{\partial R_1}{\partial \widehat{\alpha}_m^1}\frac{\partial \widehat{\alpha}_m^1}{\partial T_{\rm tr}} - \frac{R_1}{T_{\rm f}}\Big].$$

Evaluating $\frac{\partial C_0}{\partial T_{\rm tr}}$ and $\frac{\partial C_1}{\partial T_{\rm tr}}$ when $T_{\rm tr}\to 0$ we have

$$\lim_{T_{\rm tr}\to 0}\frac{\partial C_0}{\partial T_{\rm tr}} = \frac{-\beta_0}{T_{\rm f}}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}R_0\Big)}_{=0} + \beta_0\frac{(T_{\rm f}-T_{\rm se})}{T_{\rm f}} \times \sum_{m=1}^{M}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}\frac{\partial R_0}{\partial \widehat{\alpha}_m^0}\Big)}_{>0}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}\frac{\partial \widehat{\alpha}_m^0}{\partial T_{\rm tr}}\Big)}_{>0} > 0 \quad (56a)$$

$$\lim_{T_{\rm tr}\to 0}\frac{\partial C_1}{\partial T_{\rm tr}} = \frac{-\beta_1}{T_{\rm f}}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}R_1\Big)}_{=0} + \beta_1\frac{(T_{\rm f}-T_{\rm se})}{T_{\rm f}} \times \sum_{m=1}^{M}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}\frac{\partial R_1}{\partial \widehat{\alpha}_m^1}\Big)}_{>0}\underbrace{\Big(\lim_{T_{\rm tr}\to 0}\frac{\partial \widehat{\alpha}_m^1}{\partial T_{\rm tr}}\Big)}_{>0} > 0. \quad (56b)$$

Evaluating $\frac{\partial C_0}{\partial T_{\rm tr}}$ and $\frac{\partial C_1}{\partial T_{\rm tr}}$ when $T_{\rm tr}\to T_{\rm f}-T_{\rm se}$ we have

$$\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\frac{\partial C_0}{\partial T_{\rm tr}} = \lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\frac{-\beta_0 R_0}{T_{\rm f}} + \beta_0\underbrace{\Big(\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}D_{\rm d}\Big)}_{=0} \times \Big(\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\sum_{m=1}^{M}\frac{\partial R_0}{\partial \widehat{\alpha}_m^0}\frac{\partial \widehat{\alpha}_m^0}{\partial T_{\rm tr}}\Big) < 0, \quad (57a)$$

$$\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\frac{\partial C_1}{\partial T_{\rm tr}} = \lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\frac{-\beta_1 R_1}{T_{\rm f}} + \beta_1\underbrace{\Big(\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}D_{\rm d}\Big)}_{=0} \times \Big(\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\sum_{m=1}^{M}\frac{\partial R_1}{\partial \widehat{\alpha}_m^1}\frac{\partial \widehat{\alpha}_m^1}{\partial T_{\rm tr}}\Big) < 0. \quad (57b)$$

The inequalities in (56a) and (56b) show that $\lim_{T_{\rm tr}\to 0}\frac{\partial R_{\rm LB}}{\partial T_{\rm tr}} > 0$. On the other hand, the inequalities in (57a) and (57b) show that $\lim_{T_{\rm tr}\to T_{\rm f}-T_{\rm se}}\frac{\partial R_{\rm LB}}{\partial T_{\rm tr}} < 0$. Together, these indicate that the equation $\partial R_{\rm LB}/\partial T_{\rm tr}=0$ has one solution in this interval, which can be found numerically using bisection search method.

References

[1] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[2] H. Yazdani and A. Vosoughi, "On cognitive radio systems with directional antennas and imperfect spectrum sensing," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3589–3593.

[3] M. Joneidi, H. Yazdani, A. Vosoughi, and N. Rahnavard, "Source localization and tracking for dynamic radio cartography using directional antennas," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2019, pp. 1–9.

[4] Z. Rezki and M. Alouini, "Ergodic capacity of cognitive radio under imperfect channel-state information," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5, pp. 2108–2119, Jun 2012.

[5] D. Xu, Z. Feng, and P. Zhang, "On the impacts of channel estimation errors and feedback delay on the ergodic capacity for spectrum sharing cognitive radio," *Wireless Personal Communications*, vol. 72, no. 4, pp. 1875–1887, Oct 2013.

[6] H. A. Suraweera, P. J. Smith, and M. Shafi, "Capacity limits and performance analysis of cognitive radio with imperfect channel knowledge," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1811–1822, May 2010.

[7] L. Sboui, Z. Rezki, and M. Alouini, "A unified framework for the ergodic capacity of spectrum sharing cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 877–887, February 2013.

[8] P. J. Smith, P. A. Dmochowski, H. A. Suraweera, and M. Shafi, "The effects of limited channel knowledge on cognitive radio system capacity," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 927–933, Feb 2013.

[9] A. Kaushik, S. K. Sharma, S. Chatzinotas, B. Ottersten, and F. K. Jondral, "On the performance analysis of underlay cognitive radio systems: A deployment perspective," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, no. 3, pp. 273–287, Sep. 2016.

[10] S. Kashyap and N. B. Mehta, "Optimal binary power control for underlay CR with different interference constraints and impact of channel estimation errors," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3753–3764, Nov 2014.

[11] H. Yazdani and A. Vosoughi, "On the combined effect of directional antennas and imperfect spectrum sensing upon ergodic capacity of cognitive radio systems," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Oct 2017, pp. 1702–1706.

[12] H. Yazdani and A. Vosoughi, "On optimal sensing and capacity trade-off in cognitive radio systems with directional antennas," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2018, pp. 1015–1019.

[13] A. Kaushik, S. K. Sharma, S. Chatzinotas, B. Ottersten, and F. K. Jondral, "Sensing-throughput tradeoff for interweave cognitive radio system: A deployment-centric viewpoint," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3690–3702, May 2016.

[14] S. Akin and M. C. Gursoy, "Performance analysis of cognitive radio systems under QoS constraints and channel uncertainty," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 2883–2895, September 2011.

[15] S. Akin and M. C. Gursoy, "Performance analysis of cognitive radio systems with imperfect channel sensing and estimation," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1554–1566, May 2015.

[16] G. Ozcan, M. C. Gursoy, N. Tran, and J. Tang, "Energy-efficient power allocation in cognitive radio systems with imperfect spectrum sensing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3466–3481, Dec 2016.

[17] L. Zhang, Y. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel state information," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4143–4153, August 2009.

[18] F. Gao, R. Zhang, Y. Liang, and X. Wang, "Design of learning-based MIMO cognitive radio systems," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1707–1720, May 2010.

[19] R. Sarvendranath and N. B. Mehta, "Transmit antenna selection for interference-outage constrained underlay CR," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 3772–3783, Sep. 2018.

[20] S. Akin and M. C. Gursoy, "On the throughput and energy efficiency of cognitive MIMO transmissions," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3245–3260, Sep. 2013.

[21] G. A. Ropokis, M. C. Filippou, A. A. Rontogiannis, L. A. DaSilva, N. Marchetti, V. Frascolla, and P. T. Mathiopoulos, "Optimal sensing and power allocation in pilot-aided shared access systems: A BER minimization approach," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2016, pp. 1–6.

[22] W. Ouyang and X. Gong, "A 20-element cavity-backed slot electronically steerable parasitic array radiator (ESPAR) with 2-D beamsteering and minimized beam squint," *IEEE Antennas and Wireless Propagation Letters*, pp. 1–1, 2020.

[23] W. Ouyang, A. Vosoughi, and X. Gong, "A frequency-reconfigurable electronically-steerable parasitic array radiator using microstrip patch antennas," *Microwave and Optical Technology Letters*, vol. 62, no. 3, pp. 1409–1422, 2020.

[24] W. Ouyang and X. Gong, "An electronically steerable parasitic array radiator (ESPAR) using cavity-backed slot antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 4, pp. 757–761, 2019.

[25] H. Yazdani, A. Vosoughi, and N. Rahnavard, "Compressive sensing based direction-of-arrival estimation using reweighted greedy block coordinate descent algorithm for ESPAR antennas," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Oct 2017, pp. 169–173.

[26] V. W. Wong, R. Schober, and D. W. K. Ng, *Key Technologies for 5G Wireless Systems*. Cambridge University Press, 2017.

[27] M. A. Hossain, I. Bahceci, and B. A. Cetiner, "Parasitic layer-based radiation pattern reconfigurable antenna for 5G communications," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6444–6452, 2017.

[28] M. A. Almasi, H. Mehrpouyan, V. Vakilian, N. Behdad, and H. Jafarkhani, "A new reconfigurable antenna MIMO architecture for mmWave communication," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–7.

[29] D. Wilcox, E. Tsakalaki, A. Kortun, T. Ratnarajah, C. B. Papadias, and M. Sellathurai, "On spatial domain cognitive radio using single-radio parasitic antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 571–580, March 2013.

[30] C. Liu and M. Jin, "Maximum-minimum spatial spectrum detection for cognitive radio using parasitic antenna arrays," in *2014 IEEE/CIC International Conference on Communications in China (ICCC)*, Oct 2014, pp. 365–369.

[31] C. Liu, M. Li, and M. L. Jin, "Blind energy-based detection for spatial spectrum sensing," *IEEE Wireless Communications Letters*, vol. 4, no. 1, pp. 98–101, Feb 2015.

[32] H. Yazdani, A. Vosoughi, and X. Gong, "Beam selection and discrete power allocation in opportunistic cognitive radio systems with limited feedback using ESPAR antennas," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 325–339, 2020.

[33] H. Yazdani and A. Vosoughi, "On the spectrum sensing, beam selection and power allocation in cognitive radio networks using reconfigurable antennas," in *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, March 2019, pp. 1–7.

[34] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[35] M. H. Yılmaz, M. M. Abdallah, H. M. El-Sallabi, J. F. Chamberland, K. A. Qaraqe, and H. Arslan, "Joint subcarrier and antenna state selection for cognitive heterogeneous networks with reconfigurable antennas," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4015–4025, Nov 2015.

[36] F. Awin, E. Abdel-Raheem, and K. Tepe, "Blind spectrum sensing approaches for interweaved cognitive radio system: A tutorial and short course," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 238–259, Firstquarter 2019.

[37] A. Kortun, T. Ratnarajah, M. Sellathurai, Y. Liang, and Y. Zeng, "On the eigenvalue-based spectrum sensing and secondary user throughput," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1480–1486, March 2014.

[38] C. Liu, J. Wang, X. Liu, and Y. Liang, "Maximum eigenvalue-based goodness-of-fit detection for spectrum sensing in cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7747–7760, Aug 2019.

[39] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, February 2010.

[40] Y. He, T. Ratnarajah, E. H. Yousif, J. Xue, and M. Sellathurai, "Performance analysis of multi-antenna GLRT-based spectrum sensing for cognitive radio," *Signal Processing*, vol. 120, pp. 580–593, March 2016.

[41] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[42] G. Ozcan, M. C. Gursoy, and S. Gezici, "Error rate analysis of cognitive radio transmissions with imperfect channel sensing," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1642–1655, March 2014.

[43] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[44] A. Vosoughi and A. Scaglione, "On the effect of receiver estimation error upon channel mutual information," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 459–472, 2006.

[45] A. Vosoughi and Y. Jia, "How does channel estimation error affect average sum-rate in two-way amplify-and-forward relay networks?" *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1676–1687, 2012.

[46] M. Shirazi and A. Vosoughi, "On distributed estimation in hierarchical power constrained wireless sensor networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 442–459, 2020.

[47] D. P. Bertsekas and A. Scientific, *Convex optimization algorithms*. Athena Scientific Belmont, 2015.

[48] D. G. Luenberger, Y. Ye *et al.*, *Linear and nonlinear programming*, 4th ed. Springer, 2015.