Cognitive Beamforming Made Practical: Effective Interference Channel and Learning-Throughput Tradeoff

Rui Zhang, *Member, IEEE*, Feifei Gao, *Member, IEEE*, and Ying-Chang Liang, *Senior Member, IEEE*

Abstract— This paper studies the transmit strategy for a secondary link or the so-called cognitive radio (CR) link under opportunistic spectrum sharing with an existing primary radio (PR) link. It is assumed that the CR transmitter is equipped with multi-antennas, whereby transmit precoding and power control can be jointly deployed to balance between avoiding interference at the PR terminals and optimizing performance of the CR link. This operation is named as *cognitive beamforming* (CB). Unlike prior study on CB that assumes perfect knowledge of the channels over which the CR transmitter interferes with the PR terminals, this paper proposes a *practical* CB scheme utilizing a new idea of *effective interference channel* (EIC), which can be efficiently estimated at the CR transmitter from its observed PR signals. Somehow surprisingly, this paper shows that the learning-based CB scheme with the EIC improves the CR channel capacity against the conventional scheme even with the exact CR-to-PR channel knowledge, when the PR link is equipped with multi-antennas but only communicates over a subspace of the total available spatial dimensions. Moreover, this paper presents algorithms for the CR to estimate the EIC over a finite learning time. Due to channel estimation errors, the proposed CB scheme causes leakage interference at the PR terminals, which leads to an interesting *learning-throughput tradeoff* phenomenon for the CR, pertinent to its time allocation between channel learning and data transmission. This paper derives the optimal channel learning time to maximize the effective throughput of the CR link, subject to the CR transmit power constraint and the interference power constraints for the PR terminals.

Index Terms— Cognitive beamforming, cognitive radio, effective interference channel, learning-throughput tradeoff, multi-antenna systems, spectrum sharing.

I. INTRODUCTION

Cognitive radio (CR), since the name was coined by Mitola in his seminal work [1], has drawn intensive attentions from both academic and industrial communities. Generally speaking, there are three basic operation models for CRs, namely, *Interweave*, *Overlay*, and *Underlay* (see, e.g., [2] and references therein). The interweave method is also known as *opportunistic spectrum access* (OSA), originally outlined in [1] and later introduced by DARPA, where the CR is allowed to transmit over the spectrum allocated to an existing

primary radio (PR) system only when all PR transmissions are detected to be off. In contrast to interweave, the overlay and underlay methods allow the CR to transmit concurrently with PRs at the same frequency. The overlay method utilizes an interesting "cognitive relay" idea [3], [4]. For this method, the CR transmitter is assumed to know perfectly all the channels in the coexisting PR and CR links, as well as the PR messages to be sent. Thereby, the CR transmitter is able to forward PR messages to the PR receivers so as to compensate for the interference due to its own messages sent concurrently to the CR receiver. In comparison with overlay, the underlay method requires only the channel gain knowledge from the CR transmitter to the PR receivers, whereby the CR is permitted to transmit regardless of the on/off status of PR transmissions provided that its resulted signal power levels at all PR receivers are kept below some predefined threshold, also known as the *interference-temperature* constraint [5], [6]. From implementation viewpoints, interweave and underlay methods could be more favorable than overlay for practical CR systems.

In a wireless environment, channels are usually subject to space-time-frequency variation (fading) due to multipath propagation, mobility, and location-dependent shadowing. As such, *dynamic resource allocation* (DRA) becomes crucial to CRs for optimally deploying their transmit strategies, where the transmit power, bit-rate, bandwidth, and antenna beam are dynamically allocated based upon the channel state information (CSI) of the PR and CR systems (see, e.g., [7]-[14]). In this paper, we are particularly interested in the case where the CR terminal is equipped with multi-antennas so that it can deploy joint transmit precoding and power control, namely *cognitive beamforming* (CB), to effectively balance between avoiding interference at the PR terminals and optimizing performance of the CR link. In [14], various CB schemes have been proposed considering the CR transmit power constraint and a set of interference power constraints at the PR terminals, under the assumption that the CR transmitter knows perfectly all the channels over which it interferes with PR terminals. In this work, however, we propose a *practical* CB scheme, which does not require any prior knowledge of the CR-to-PR channels. Instead, by exploiting the time-division-duplex (TDD) operation mode of the PR link and the channel reciprocities between the CR and PR terminals, the proposed CB scheme utilizes a new idea so-called *effective interference channel* (EIC), which can be efficiently estimated at the CR terminal via periodically observing the PR transmissions. Thereby, the proposed learning-based CB scheme eliminates

the overhead for PR terminals to estimate the CR-to-PR channels and then feed them back to the CR, and thus makes the CB implementable in practical systems.

Furthermore, the proposed learning-based CB scheme with the EIC creates a new operation model for CRs, where the CR is able to transmit with PRs at the same time and frequency over the detected available spatial dimensions, thus named as *opportunistic spatial sharing* (OSS). On the one hand, OSS, like the underlay method, utilizes the spectrum more efficiently than the interweave method by allowing the CR to transmit concurrently with PRs. On the other hand, OSS can further improve the CR transmit spectral efficiency over the underlay method by exploiting additional side information on PR transmissions, which is extractable from the observed EIC (more details will be given later in this paper). Therefore, OSS is a more superior operation model for CRs than both underlay and interweave methods in terms of the spectrum utilization efficiency.

The main results of this paper constitute two parts, which are summarized as follows:

- First, we consider the ideal case where the CR's estimate on the EIC is *perfect* or noiseless. For this case, we derive the conditions under which the EIC is sufficient for the proposed CB scheme to cause no adverse effects on the concurrent PR transmissions. In addition, we show that when the PR link is equipped with multi-antennas but only communicates over a subspace of the total available spatial dimensions, the learning-based CB scheme with the EIC leads to a capacity gain over the conventional zero-forcing (ZF) scheme [14] even with the exact CR-to-PR channel knowledge, via exploiting side information on PR transmit dimensions extracted from the EIC.
- Second, we consider the practical case with *imperfect* estimation of EIC due to finite learning time. We propose a *two-phase* protocol for CRs to implement learning-based CB. The first phase is for the CR to observe the PR signals and estimate the EIC, while the second phase is for the CR to transmit data with CB designed via the estimated EIC. We present two algorithms for CRs to estimate the EIC, under different assumptions on the availability of the noise power knowledge at the CR terminal. Furthermore, due to imperfect channel estimation, the proposed CB scheme results in leakage interference at the PR terminals, which leads to an interesting *learning-throughput tradeoff*, i.e., different choices of time allocation between CR's channel learning and data transmission correspond to different tradeoffs between PR transmission protection and CR throughput maximization. We formulate the problem to determine the optimal time allocation for estimating the EIC to maximize the effective throughput of the CR link, subject to the CR transmit power constraint and the interference power constraints at the PR terminals; and derive the solution via applying convex optimization techniques.

The rest of this paper is organized as follows. Section II presents the CR system model. Section III introduces the idea of EIC. Section IV studies the CB design based on the EIC un-
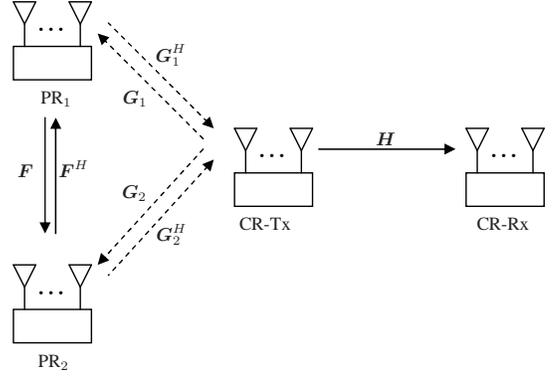

Fig. 1. Spectrum sharing between a CR link and a PR link.

der perfect channel learning. Section V considers the case with imperfect channel learning, presents algorithms for estimating the EIC, and studies the learning-throughput tradeoff for the CR link. Section VI presents numerical results to corroborate the proposed studies. Finally, Section VII concludes the paper.

Notation: Scalar is denoted by lower-case letter, e.g., x, and bold-face lower-case letter is used for vector, e.g., $\boldsymbol{x}$, and bold-face upper-case letter is for matrix, e.g., $\boldsymbol{X}$. For a matrix $\boldsymbol{S}$, $\text{Tr}(\boldsymbol{S})$, $\text{Rank}(\boldsymbol{S})$, $|\boldsymbol{S}|$, $\boldsymbol{S}^{-1}$, $\boldsymbol{S}^{\dagger}$, $\boldsymbol{S}^{T}$, and $\boldsymbol{S}^{H}$ denote its trace, rank, determinant, inverse, pseudo inverse, transpose, and conjugate transpose, respectively. $\text{Diag}(x_1, \ldots, x_M)$ denotes a $M \times M$ diagonal matrix with diagonal elements given by $x_1, \ldots, x_M$. For a matrix $\boldsymbol{M}$, $\lambda_{\max}(\boldsymbol{M})$ and $\lambda_{\min}(\boldsymbol{M})$ denote the maximum and minimum eigenvalues of $\boldsymbol{M}$, respectively. $\boldsymbol{I}$ and $\boldsymbol{0}$ denote the identity matrix and the all-zero matrix, respectively, with proper dimensions. For a positive semi-definite matrix $\boldsymbol{S}$, denoted by $\boldsymbol{S} \succeq \boldsymbol{0}$, $\boldsymbol{S}^{1/2}$ denotes a square-root matrix of $\boldsymbol{S}$, i.e., $\boldsymbol{S}^{1/2}(\boldsymbol{S}^{1/2})^H = \boldsymbol{S}$, which is assumed to be obtained from the eigenvalue decomposition (EVD) of $\boldsymbol{S}$: If the EVD of $\boldsymbol{S}$ is expressed as $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$, then $\boldsymbol{S}^{1/2} = \boldsymbol{U}\boldsymbol{\Sigma}^{1/2}$. $\|\boldsymbol{x}\|$ denotes the Euclidean norm of a complex vector $\boldsymbol{x}$. $\mathbb{C}^{x \times y}$ denotes the space of $x \times y$ matrices with complex entries. The distribution of a circular symmetric complex Gaussian (CSCG) vector with mean vector $\boldsymbol{x}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{x}, \boldsymbol{\Sigma})$, and $\sim$ stands for "distributed as". $\mathbb{E}[\cdot]$ denotes the statistical expectation. $\text{Prob}\{\cdot\}$ denotes the probability. $\max(x, y)$ and $\min(x, y)$ denote the maximum and the minimum of two real numbers, x and y, respectively. For a real number a, $(a)^+ = \max(0, a)$.

II. SYSTEM MODEL

For the purpose of exposition, in this paper we consider a simplified CR system as shown in Fig. 1, where a single CR link consisting of one CR transmitter (CR-Tx) and one CR receiver (CR-Rx) coexists with a single PR link consisting of two terminals denoted by PR_1 and PR_2, respectively. The number of antennas equipped at CR-Tx, CR-Rx, PR_1, and PR_2 are denoted by M_t, M_r, M_1, and M_2, respectively. It is assumed that $M_t > 1$, while M_r, M_1, and M_2 can be any positive integers. For the PR link, it is assumed that PR_1 and PR_2 operate in a TDD mode over a narrow-band

flat-fading channel. Furthermore, reciprocity is assumed for the channels between PR_1 and PR_2, i.e., if the channel from PR_1 to PR_2 is denoted by $\boldsymbol{F} \in \mathbb{C}^{M_2 \times M_1}$, then the channel from PR_2 to PR_1 becomes $\boldsymbol{F}^H$.[1] Without loss of generality (W.l.o.g.), the transmit beamforming matrix for PR_j, $j = 1, 2$, is denoted by $\boldsymbol{A}_j \in \mathbb{C}^{M_j \times d_j}$, with d_j denoting the corresponding number of transmit data streams, $1 \le d_j \le M_j$. The transmit covariance matrix for PR_j is then defined as $\boldsymbol{S}_j \triangleq \boldsymbol{A}_j \boldsymbol{A}_j^H$. We assume that $\boldsymbol{A}_j$ is a full-rank matrix and thus $\text{Rank}(\boldsymbol{S}_j) = d_j$. Furthermore, define $\boldsymbol{B}_1 \in \mathbb{C}^{d_2 \times M_1}$ as the receive beamforming matrix for PR_1 and $\boldsymbol{B}_2 \in \mathbb{C}^{d_1 \times M_2}$ for PR_2. Both $\boldsymbol{B}_j$'s are assumed to be full-rank. It is also assumed that PR_1 and PR_2 are both oblivious to the CR, and treat the interference from the CR as additional noise.

The CR is assumed to transmit over the same frequency band of the PR, and thus it needs to protect any active PR transmissions by limiting its resulted interference power levels at both PR_j's to be below some prescribed threshold (to be specified later in Section V). Let $\boldsymbol{H} \in \mathbb{C}^{M_r \times M_t}$ denote the CR channel, and $\boldsymbol{G}_j \in \mathbb{C}^{M_j \times M_t}$ denote the interference channel from CR-Tx to PR_j, $j = 1, 2$. Let the transmit beamforming matrix of CR-Tx be denoted by a full-rank matrix $\boldsymbol{A}_{\text{CR}} \in \mathbb{C}^{M_t \times d_{\text{CR}}}$, where $d_{\text{CR}} \le M_t$ and $d_{\text{CR}} = \text{Rank}(\boldsymbol{S}_{\text{CR}})$, with $\boldsymbol{S}_{\text{CR}}$ denoting the transmit covariance matrix of CR-Tx, i.e., $\boldsymbol{S}_{\text{CR}} \triangleq \boldsymbol{A}_{\text{CR}} \boldsymbol{A}_{\text{CR}}^H$. Note that in Fig. 1, the channels from PR_j's to CR-Rx are not shown, while similarly as for the PR terminals, it is assumed that any interference from PR_j's over these channels is treated as additional noise at CR-Rx.

In [14], various CB designs in terms of the CR transmit covariance matrix, $\boldsymbol{S}_{\text{CR}}$, have been studied for a similar system setup like that in Fig. 1, where the CR channel transmit rate is maximized under the CR transmit power constraint and a set of interference power constraints for the PR terminals. The CB designs in [14] assume that in Fig. 1, CR-Tx has perfect knowledge of $\boldsymbol{H}$, $\boldsymbol{G}_1$, and $\boldsymbol{G}_2$. In this work, however, we remove the assumption on any prior knowledge of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ at CR-Tx for the design of CB, since in practice the CR and PR systems usually belong to different operators, and it is thus difficult to require PRs to estimate CR-to-PR channels and then feed them back to the CR. As such, the most feasible method for the CR to learn some knowledge of CR-to-PR channels is to observe the PR signals propagating through PR-to-CR channels and then apply the channel reciprocities between CR-Tx and PR_j's. Thus, in this paper we propose a *learning-based* CR transmit strategy, where the CR first observes the received PR signals to extract CR-to-PR channel knowledge, and then designs the CB based upon the obtained channel knowledge. However, there are several issues related to this approach, which are pointed out as follows:

- What CR-Tx can possibly estimate are indeed the "effective" channels, $\boldsymbol{G}_j^H \boldsymbol{A}_j$, from PR_j, $j = 1, 2$, instead of the actual channels, $\boldsymbol{G}_j$'s, if the PR transmit beamforming matrices, $\boldsymbol{A}_j$'s, are not known at CR-Tx.
- The proposed CB schemes in [14] require that the channels associated with $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ be separately estimated

at CR-Tx. As such, CR-Tx needs to synchronize with PR TDD transmissions, which requires knowledge of the exact time instants over each transmit direction between PR_1 and PR_2.
- If CR-Tx designs $\boldsymbol{S}_{\text{CR}}$ based on the effective channels, $\boldsymbol{G}_j^H \boldsymbol{A}_j$'s, it is unclear whether the effect of its resulted interference at PR_j's can be properly controlled since the signals from CR-Tx interferer with PR_j via the equivalent channel, $\boldsymbol{B}_j \boldsymbol{G}_j$, which can be different from $\boldsymbol{A}_j^H \boldsymbol{G}_j$ if the PR receive beamforming matrix $\boldsymbol{B}_j$ differs from $\boldsymbol{A}_j$.

Therefore, to make the learning-based CB feasible for practical systems, the above issues need to be carefully addressed, without critical assumptions or prior knowledge on PR signal processing procedures. In this paper, we propose an effective solution to resolve the aforementioned issues, while utilizes a new idea, namely *effective interference channel* (EIC), as will be presented next.

III. EFFECTIVE INTERFERENCE CHANNEL

For the learning-based CB scheme, suppose that prior to data transmission, CR-Tx first listens to the frequency band of interest for PR transmissions over N symbol periods. The received baseband signals can then be represented as

$$\boldsymbol{y}(n) = \boldsymbol{G}_j^H \boldsymbol{A}_j \boldsymbol{t}_j(n) + \boldsymbol{z}(n), \ n = 1, \ldots, N \qquad (1)$$

where $j = 1$ if $n \in \mathcal{N}_1$, and $j = 2$ if $n \in \mathcal{N}_2$, with $\mathcal{N}_1, \mathcal{N}_2 \subseteq \{1, \ldots, N\}$ denoting the time instants when PR_1 transmits to PR_2 and PR_2 transmits to PR_1, respectively, and $\mathcal{N}_1 \cap \mathcal{N}_2 = \varnothing$ due to the assumed TDD mode; $\boldsymbol{t}_j(n)$'s are the encoded signals (prior to power control and precoding) for the corresponding PR_j, and for convenience it is assumed that $\boldsymbol{t}_j(n)$'s are independent over n's and $\mathbb{E}[\boldsymbol{t}_j(n)(\boldsymbol{t}_j(n))^H] = \boldsymbol{I}_{d_j \times d_j}$, $j = 1, 2$; $\boldsymbol{z}(n)$'s are the additive noises assumed to be independent CSCG random vectors with zero-mean and covariance matrix denoted by $\rho_0 \boldsymbol{I}_{M_t \times M_t}$. Denote the cardinality of the set $\mathcal{N}_j$ as $|\mathcal{N}_j|$. It is reasonable to assume that PR_j will transmit, with a constant probability $\alpha_j < 1$, during a certain time period. Mathematically, we may use $\mathbb{E}\left[\frac{|\mathcal{N}_j|}{N} \Big| N\right] = \alpha_j$ or $\mathbb{E}\left[\frac{|\mathcal{N}_j|}{N}\right] = \alpha_j$. Note that $\alpha_1 + \alpha_2 \le 1$, where a strict inequality occurs when there are guard (silent) intervals between alternate PR TDD transmissions. Also note that if $\alpha_1 = \alpha_2 = 0$, there will be no active PR transmissions in the observed frequency band.

Define $\boldsymbol{s}_j(n)$ as $q_j(n)\boldsymbol{t}_j(n)$, where $q_j(n) = 1$, if $n \in \mathcal{N}_j$ and $q_j(n) = 0$ otherwise. Obviously, $q_j(n)$'s are random variables with $\mathbb{E}[q_j(n)] = \alpha_j$. Meanwhile, $q_1(n)$ and $q_2(n)$ are related by $q_1(n)q_2(n) = 0, \forall n$. Thus, we have $\mathbb{E}[\boldsymbol{s}_j(n)(\boldsymbol{s}_j(n))^H] = \alpha_j \boldsymbol{I}$, $j = 1, 2$, but $\mathbb{E}[\boldsymbol{s}_1(n)(\boldsymbol{s}_2(n))^H] = \boldsymbol{0}$. The signal model in (1) can then be equivalently rewritten as

$$\boldsymbol{y}(n) = \boldsymbol{\mathcal{A}}\boldsymbol{s}(n) + \boldsymbol{z}(n), \ n = 1, \ldots, N \qquad (2)$$

where $\boldsymbol{\mathcal{A}} = [\boldsymbol{G}_1^H \boldsymbol{A}_1, \boldsymbol{G}_2^H \boldsymbol{A}_2]$ and $\boldsymbol{s}(n) = [(\boldsymbol{s}_1(n))^T, (\boldsymbol{s}_2(n))^T]^T$. The covariance matrix of the received signals at CR-Tx is then defined as

$$\boldsymbol{Q}_y = \mathbb{E}[\boldsymbol{y}(n)(\boldsymbol{y}(n))^H] = \boldsymbol{Q}_s + \rho_0 \boldsymbol{I} \qquad (3)$$

---

[1]The results of this paper hold similarly for the case where $\boldsymbol{F}^T$ instead of $\boldsymbol{F}^H$ is used to represent the reverse channel of $\boldsymbol{F}$.

where

$$\boldsymbol{Q}_s \triangleq \alpha_1 \boldsymbol{G}_1^H \boldsymbol{S}_1 \boldsymbol{G}_1 + \alpha_2 \boldsymbol{G}_2^H \boldsymbol{S}_2 \boldsymbol{G}_2 \qquad (4)$$

denotes the covariance matrix due to only the signals from PR_j's.

Practically, only the sample covariance matrix can be obtained at CR-Tx, which is expressed as

$$\hat{\boldsymbol{Q}}_y = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}(n)(\boldsymbol{y}(n))^H. \qquad (5)$$

From law of large number (LLN), it is easy to verify that $\hat{\boldsymbol{Q}}_y \to \boldsymbol{Q}_s + \rho_0 \boldsymbol{I}$ with probability one as $N \to \infty$, while for finite values of N, $\boldsymbol{Q}_s$ can only be estimated from $\hat{\boldsymbol{Q}}_y$.[2] Denote $\hat{\boldsymbol{Q}}_s$ as the estimate of $\boldsymbol{Q}_s$ from $\hat{\boldsymbol{Q}}_y$. Note that $\hat{\boldsymbol{Q}}_s$ should be a covariance matrix and hence $\hat{\boldsymbol{Q}}_s \succeq \boldsymbol{0}$ and $\hat{\boldsymbol{Q}}_s^H = \hat{\boldsymbol{Q}}_s$. Next, we denote the aggregate "effective" channel from both PR_j's to CR-Tx as

$$\boldsymbol{G}_{\mathrm{eff}}^H = \hat{\boldsymbol{Q}}_s^{1/2} \qquad (6)$$

while under the assumption of channel reciprocity, we denote the *effective interference channel* (EIC) from CR-Tx to both PR_j's as $\boldsymbol{G}_{\mathrm{eff}}$.

In the rest of this paper, CB schemes based on the EIC instead of the actual CR-to-PR channels will be studied. Note that with the EIC, the first two items of implementation issues raised in Section II are resolved. The first issue is resolved since the EIC is defined over the effective channels from PR_j's to CR-Tx instead of the actual channels, while the second issue is resolved since the EIC does not attempt to separate the two channels from PR_j's to CR-Tx, and thus synchronization for CR-Tx with each transmit direction between PR_1 and PR_2 is no longer required. However, we still need to address the third issue on analyzing the effect of the CR's interference on the PR transmissions with the EIC-based CB design. We will first address this issue for the ideal case where the estimation of $\boldsymbol{G}_{\mathrm{eff}}$ is perfect or noiseless in Section IV, in order to gain some insights into this problem. Then, we will study this problem for the more practical case where $\boldsymbol{G}_{\mathrm{eff}}$ is imperfectly estimated due to finite values of N in Section V.

IV. PERFECT CHANNEL LEARNING

In this section, we design the CR transmit covariance matrix, $\boldsymbol{S}_{\mathrm{CR}}$, in terms of the equivalent transmit beamforming matrix, $\boldsymbol{A}_{\mathrm{CR}}$, which contains information of both transmit precoding and power allocation at CR-Tx, under perfect learning of the EIC, i.e., the noise effect on estimating $\boldsymbol{Q}_s$ from $\hat{\boldsymbol{Q}}_y$ is completely removed. For this case, $\hat{\boldsymbol{Q}}_s = \boldsymbol{Q}_s$ in (6), and from (4) it follows that the EIC can now be expressed as

$$\boldsymbol{G}_{\mathrm{eff}} = \left(\left(\alpha_1 \boldsymbol{G}_1^H \boldsymbol{S}_1 \boldsymbol{G}_1 + \alpha_2 \boldsymbol{G}_2^H \boldsymbol{S}_2 \boldsymbol{G}_2 \right)^{1/2} \right)^H. \qquad (7)$$

From (7), due to independence of the channels $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$, it follows that $d_{\mathrm{eff}} = \mathrm{Rank}(\boldsymbol{G}_{\mathrm{eff}}) = \min(d_1 + d_2, M_t)$. Thus, if the number of antennas at CR-Tx, M_t, is strictly greater than the total number of transmit data streams between PR_1

and PR_2, $d_1 + d_2$, then the EIC-based CB design will have at most $M_t - (d_1 + d_2)$ number of spatial dimensions or degrees of freedom (DoF) [15] for transmission, where all these dimensions lie in the null space of $\boldsymbol{G}_{\mathrm{eff}}$. Based on this observation, we obtain the following proposition:

Proposition 4.1: Under perfect learning of the EIC, if the conditions $\boldsymbol{A}_j^H \boldsymbol{G}_j \sqsupseteq \boldsymbol{B}_j \boldsymbol{G}_j, j = 1, 2$ hold,[3] then the EIC-based CB design satisfying the constraint $\boldsymbol{G}_{\mathrm{eff}} \boldsymbol{A}_{\mathrm{CR}} = \boldsymbol{0}$ will have no adverse effects on PR transmissions, i.e., $\boldsymbol{B}_j \boldsymbol{G}_j \boldsymbol{A}_{\mathrm{CR}} = \boldsymbol{0}, j = 1, 2$.

The conditions in the above proposition can also be expressed as $\mathrm{Span}(\boldsymbol{A}_j^H \boldsymbol{G}_j) \supseteq \mathrm{Span}(\boldsymbol{B}_j \boldsymbol{G}_j), j = 1, 2$, where $\mathrm{Span}(\boldsymbol{X})$ denotes the subspace spanned by the row vectors in $\boldsymbol{X}$. Intuitively speaking, these conditions hold when the transmit signal space of PR_j after propagating through the PR-to-CR channel $\boldsymbol{G}_j^H$, i.e., $\boldsymbol{G}_j^H \boldsymbol{A}_j$, if being reversed (conjugate transposed), will subsume the equivalent channel from CR-Tx to PR_j, $\boldsymbol{B}_j \boldsymbol{G}_j$, as a subspace, for both $j = 1, 2$. Note that $\boldsymbol{A}_j^H \boldsymbol{G}_j$ and $\boldsymbol{B}_j \boldsymbol{G}_j$ may not have the same column size, and $\boldsymbol{A}_j^H$ and $\boldsymbol{B}_j$ may differ from each other for any $j = 1, 2$. As such, the validity of these conditions needs to be examined for practical systems. Thus, before we proceed to the proof of Proposition 4.1, we present two typical examples of multi-antenna transmission schemes for the PR link as follows, for both of which the conditions $\boldsymbol{A}_j^H \boldsymbol{G}_j \sqsupseteq \boldsymbol{B}_j \boldsymbol{G}_j, j = 1, 2$ are usually satisfied.[4]

Example 4.1: Spatial Multiplexing: When the PR CSI is unknown at transmitter but known at receiver, the spatial multiplexing mode is usually adopted to assign equal powers and rates to transmit antennas (e.g., the V-BLAST scheme [15]). For this case, the transmit covariance matrix at PR_j, $j = 1, 2$, reduces to $\boldsymbol{S}_j = \frac{P_j}{M_j} \boldsymbol{I}_{M_j \times M_j}$, with P_j denoting the transmit power of PR_j. Thus, $d_j = M_j$, and $\boldsymbol{A}_j$'s are both scaled identity matrices. It then follows that $\boldsymbol{A}_j^H \boldsymbol{G}_j \sqsupseteq \boldsymbol{B}_j \boldsymbol{G}_j, j = 1, 2$, holds regardless of $\boldsymbol{B}_j$.

Example 4.2: Eigenmode Transmission: When the PR CSI is known at both transmitter and receiver, which is usually a valid assumption for the TDD mode, the eigenmode transmission mode is usually adopted to decompose the multi-antenna PR channel into parallel scalar channels [15]. For this case, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are designed based on the singular-value decomposition (SVD) of $\boldsymbol{F}$ and $\boldsymbol{F}^H$, respectively. Let the SVD of $\boldsymbol{F}$ be $\boldsymbol{U}_F \boldsymbol{\Sigma}_F \boldsymbol{V}_F^H$. It then follows that $\boldsymbol{A}_1 = \boldsymbol{V}_{F(1)} \boldsymbol{\Lambda}_1^{1/2}$, $\boldsymbol{B}_1 = \boldsymbol{V}_{F(2)}^H$, $\boldsymbol{A}_2 = \boldsymbol{U}_{F(2)} \boldsymbol{\Lambda}_2^{1/2}$, and $\boldsymbol{B}_2 = \boldsymbol{U}_{F(1)}^H$, where $\boldsymbol{\Lambda}_j = \mathrm{Diag}(\lambda_{j,1}, \ldots, \lambda_{j,d_j})$ is a positive diagonal matrix with $\lambda_{j,i}, i = 1, \ldots, d_j$, denoting the power allocation over the ith transmit data stream, and $\boldsymbol{V}_{F(j)}$ ($\boldsymbol{U}_{F(j)}$) consists of the first d_j columns in $\boldsymbol{V}_F$ ($\boldsymbol{U}_F$). Note that in this case $d_j \leq \min(M_1, M_2)$. If it is true that $d_1 = d_2$, i.e., both transmit directions between PR_1 and PR_2 have the same number of data streams, then it follows that $\boldsymbol{A}_j^H \boldsymbol{G}_j \sqsupseteq \boldsymbol{B}_j \boldsymbol{G}_j$ holds for both $j = 1, 2$. Note that a valid special case here is the "beamforming mode" [15] with $d_1 = d_2 = 1$.

---

[2]Algorithms for such an estimation are discussed in Section V-A.

[3]$\boldsymbol{X} \sqsupseteq \boldsymbol{Y}$ means that for two given matrices with the same collum size, $\boldsymbol{X}$ and $\boldsymbol{Y}$, if $\boldsymbol{X} \boldsymbol{e} = \boldsymbol{0}$ for any arbitrary vector $\boldsymbol{e}$, then $\boldsymbol{Y} \boldsymbol{e} = \boldsymbol{0}$ must hold.

[4]Note that when the conditions in Proposition 4.1 are not satisfied, the proposed CB scheme will cause certain performance loss to PR transmissions even under perfect channel learning.

Next, we present the proof of Proposition 4.1 as follows:

Proof: First, with perfectly known G_{eff}, $G_{\text{eff}} \sqsupseteq A_j^H G_j$ is true for $j = 1, 2$. This can be shown as follows given any arbitrary vector e: $G_{\text{eff}}e = 0 \overset{(a)}{\Rightarrow} (\hat{Q}_s^{1/2})^H e = 0 \overset{(b)}{\Rightarrow} (Q_s^{1/2})^H e = 0 \Rightarrow e^H Q_s e = 0 \overset{(c)}{\Rightarrow} \|A_j^H G_j e\|^2 = 0, j = 1, 2 \Rightarrow A_j^H G_j e = 0, j = 1, 2$, where (a) is from (6), (b) is due to $\hat{Q}_s = Q_s$, and (c) is from (4). Since for arbitrary matrices X, Y, and Z, $X \sqsupseteq Y$ and $Y \sqsupseteq Z$ imply that $X \sqsupseteq Z$, from $G_{\text{eff}} \sqsupseteq A_j^H G_j$ (shown above) and $A_j^H G_j \sqsupseteq B_j G_j$ (given in Proposition 4.1) it follows that $G_{\text{eff}} \sqsupseteq B_j G_j$, $j = 1, 2$. Therefore, if the constraint $G_{\text{eff}} A_{\text{CR}} = 0$ is satisfied, it follows that $B_j G_j A_{\text{CR}} = 0, j = 1, 2$, i.e., the interference from CR-Tx at PR_j lies in the null space of the corresponding receiver beamforming matrix B_j, and thus has no adverse effects. ∎

From Proposition 4.1, it is known that if the given conditions are satisfied, it is sufficient for us to design A_{CR} subject to the constraint $G_{\text{eff}} A_{\text{CR}} = 0$, in order to remove the effects of the CR signals on PR transmissions. Let the EVD of Q_s be represented as $Q_s = V \Sigma V^H$, where $V \in \mathbb{C}^{M_t \times d_{\text{eff}}}$ and Σ is a positive $d_{\text{eff}} \times d_{\text{eff}}$ diagonal matrix. Note that $\text{Rank}(Q_s) = d_{\text{eff}}$. From (6), G_{eff}^H can then be represented as $G_{\text{eff}}^H = V \Sigma^{1/2}$. Define the projection matrix related to V as $P_V \triangleq I - VV^H = UU^H$, where $U \in \mathbb{C}^{M_t \times (M_t - d_{\text{eff}})}$ satisfies $V^H U = 0$. We are now ready to present the general form of A_{CR} satisfying the constraint $G_{\text{eff}} A_{\text{CR}} = 0$ as [16]

$$A_{\text{CR}} = U C_{\text{CR}}^{1/2} \tag{8}$$

where $C_{\text{CR}}^{1/2} \in \mathbb{C}^{(M_t - d_{\text{eff}}) \times d_{\text{CR}}}$ with d_{CR} denoting the number of transmit data streams of the CR, and $C_{\text{CR}} \in \mathbb{C}^{(M_t - d_{\text{eff}}) \times (M_t - d_{\text{eff}})}$ satisfies that $C_{\text{CR}} \succcurlyeq 0$ and $\text{Tr}(C_{\text{CR}}) = \text{Tr}(S_{\text{CR}}) \leq P_{\text{CR}}$, with P_{CR} denoting the transmit power constraint of CR-Tx. From (8), it follows that designing the transmit beamforming matrix A_{CR} for the CR channel becomes equivalent to designing the transmit covariance matrix C_{CR} for an auxiliary multi-antenna channel, HU, subject to transmit power constraint, $\text{Tr}(C_{\text{CR}}) \leq P_{\text{CR}}$. This observation simplifies the design for the remaining part in A_{CR}, i.e., C_{CR}, since existing solutions (see, e.g., [15] and references therein) are available for this well-studied precoder design problem.

At last, we demonstrate an interesting property for the proposed CB scheme given in (8), when the conditions given in Proposition 4.1 are satisfied, and furthermore, PR_1 and/or PR_2 have multi-antennas but transmit only over a subspace of the available spatial dimensions, i.e., $d_j < \min(M_1, M_2), j = 1, 2$. For this case, we will show that the proposed scheme in (8) with the EIC G_{eff} can be superior over the conventional "projected-channel SVD (P-SVD)" scheme proposed in [14] with the actual CR-to-PR channels G_1 and G_2, in terms of the achievable DoF for CR transmission. At a first glance, this result is counter-intuitive since G_{eff} contains only partial information on G_j's. The key observation here is that G_{eff} contains information on $A_j^H G_j$'s, which also exhibit side information on $B_j G_j$'s via the conditions, $A_j^H G_j \sqsupseteq B_j G_j, j = 1, 2$, given in Proposition 4.1, while $B_j G_j$'s are assumed to be unknown for the P-SVD scheme. More specifically, for the proposed scheme, the DoF is given as d_{CR}, which can
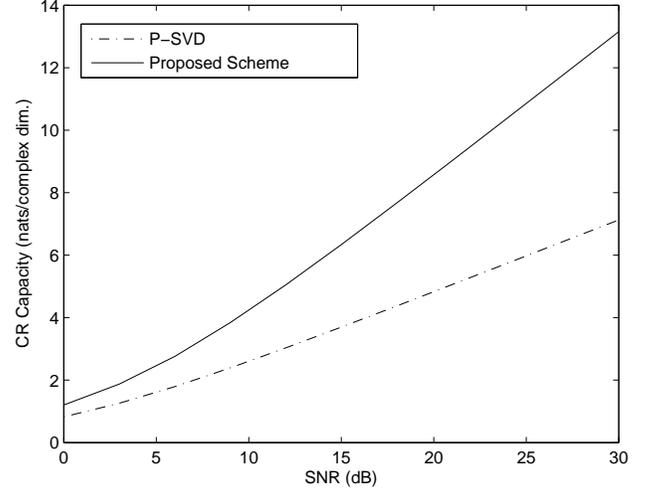

Fig. 2. CR capacity comparison for the proposed CB scheme and the P-SVD scheme in [14].

be shown to be upper-bounded by $\min(M_t - d_{\text{eff}}, M_r) = \min((M_t - d_1 - d_2)^+, M_r)$. In comparison with the proposed scheme, the P-SVD scheme with perfect knowledge of G_1 and G_2 removes the interference (thus having no effects on PR transmissions like the proposed scheme) at both PR_1 and PR_2 via transmitting only over the subspace of H that is orthogonal to both G_1 and G_2, thus resulting in the DoF to be at most $\min((M_t - M_1 - M_2)^+, M_r)$. Therefore, the proposed scheme can have a strictly positive DoF even when $M_1 + M_2 \geq M_t$, provided that $d_1 + d_2 < M_t$, i.e., the total number of antennas of PR_j's is no smaller than M_t, while the total number of data streams over both transmit directions between PR_1 and PR_2 is smaller than M_t, while the P-SVD scheme has a zero DoF in this case since $M_t \leq M_1 + M_2$. In most practical cases, we have $d_j \leq \min(M_1, M_2), j = 1, 2$. It thus follows that $(d_1 + d_2) \leq (M_1 + M_2)$ and thus the DoF gain of the proposed scheme against the P-SVD scheme, $\min((M_t - d_1 - d_2)^+, M_r) - \min((M_t - M_1 - M_2)^+, M_r)$, is always non-negative, while the maximum DoF gain occurs when $d_1 = d_2 = 0$, i.e., when the PR link switches off transmissions. Note that this DoF gain is achieved by the CR via exploiting side information on PR transmit (on/off) status or signal dimensions extracted from the observed EIC. This also justifies our previous claim that the OSS operation mode with learning-based CB is potentially more spectral efficient than the conventional interweave and underlay methods.

Example 4.3: The capacity gain of the proposed scheme in (8) over the P-SVD scheme in [14], as above discussed, is shown in Fig. 2 for a PR link with $M_1 = M_2 = 2$, $d_1 = d_2 = 1$ (i.e., beamforming mode corresponding to the largest channel singular value in Example 4.2 is used), and a CR link with $M_t = 5$ and $M_r = 3$. All the channels involved are assumed to have the standard Rayleigh-fading distribution, i.e., each element of the channel matrix is independent CSCG random variable $\sim \mathcal{CN}(0, 1)$. For simplicity, it is assumed that the interference due to PR transmissions at CR-Rx is included in the additive noise, which is assumed to be $\mathcal{CN}(0, \rho_1 I)$.
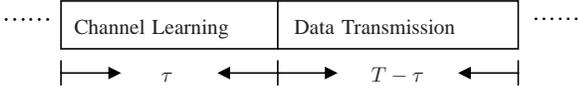
Fig. 3. Two-phase protocol for CR block-based transmission.

The signal-to-noise ratio (SNR) in this case is thus defined as P_{CR}/ρ_1. The DoF can be visually seen in the figure to be proportional to the asymptotic ratio between the capacity value over the log-SNR value as SNR goes to infinity [15]. It is observed that the DoF for the proposed scheme is approximately three times of that for the P-SVD scheme in this case, since $\min((M_t - d_1 - d_2)^+, M_r)/\min((M_t - M_1 - M_2)^+, M_r) = 3/1 = 3$.

V. IMPERFECT CHANNEL LEARNING

In the previous section, CB designs have been studied under the assumption that the EIC G_{eff} is perfectly estimated at CR-Tx. In this section, we will study the effect of imperfect estimation of G_{eff} due to finite sample size N on the performance of the proposed CB scheme. First, consider the following *two-phase* protocol for CRs to support the learning-based CB scheme as shown in Fig. 3, where each block transmission of the CR with duration T is divided into two consecutive sub-blocks. During the first sub-block of duration τ, the CR observes the PR transmissions and estimates G_{eff}; during the second sub-block of duration $T - \tau$, the CR transmits data with the CB design based on the estimated G_{eff} (following the same procedure as in Section IV, but with the estimated G_{eff} instead of its true value). Note that T needs to be chosen such that, on the one hand, to be sufficiently small compared with the channel coherence time in order to maintain constant channels during each transmission block, and on the other hand, to be as large as possible compared with the inverse of the channel bandwidth in order to make T constitute a large number of symbols to reduce the overhead of channel learning. In this paper, we assume that T is preselected as a fixed value. For a given T, intuitively, a larger value of τ is desirable from the perspective of estimating G_{eff}, while a smaller τ is favorable in terms of the effective CR link throughput that is proportional to $(T - \tau)/T$. Therefore, there exists a general *learning-throughput tradeoff* for the proposed CB scheme,[5] where different choices for the value τ lead to different tradeoffs between PR transmission protection and CR throughput maximization.

The rest of this section is organized as follows. We first present two practical algorithms for CRs to estimate G_{eff} with a finite N in Section V-A. Then, in Section V-B we analyze the so-called "effective leakage interference" at PR terminals for the proposed CB scheme due to estimation errors in G_{eff}. At last, in Section V-C we characterize the learning-throughput tradeoff for CRs by finding the optimal value of τ to maximize

[5]Note that the learning-throughput tradeoff includes the sensing-throughput tradeoff studied in [17] as a special case since CR channel sensing to detect PR's on/off status [17] can be considered as a "hard" version of CR channel learning proposed in this paper.

the CR link effective throughput, under the given T, P_{CR}, and maximum tolerable leakage interference power level at PR_j's.

A. Estimation of G_{eff}

From (6), it is known that G_{eff} depends solely on $\hat{Q}_s$, which is the estimate of the received PR signal covariance matrix Q_s defined in (4). Thus, in this subsection, we present two algorithms to obtain $\hat{Q}_s$ from the received sample covariance matrix $\hat{Q}_y$ given in (5). Denote the EVD of $\hat{Q}_y$ as

$$\hat{Q}_y = \hat{T}_y \hat{\Lambda}_y \hat{T}_y^H \qquad (9)$$

where $\hat{\Lambda}_y = \mathrm{Diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_{M_t})$ is a positive diagonal matrix whose diagonal elements are the eigenvalues of $\hat{Q}_y$. W.l.o.g., we assume that $\hat{\lambda}_i$'s, $i = 1, \ldots, M_t$, are arranged in a decreasing order. We then obtain $\hat{Q}_s$ from $\hat{Q}_y$ based on the standard maximum likelihood (ML) criterion, for the following two cases:

1) Known noise power ρ_0: In the case where the noise power, ρ_0, is assumed to be known at CR-Tx prior to channel learning, it follows from [18] that the ML estimate of Q_s is obtained as

$$\hat{Q}_s = \hat{T}_y \mathrm{Diag}\left((\hat{\lambda}_1 - \rho_0)^+, \ldots, (\hat{\lambda}_{M_t} - \rho_0)^+\right) \hat{T}_y^H. \quad (10)$$

The rank of $\hat{Q}_s$, or the estimate of d_{eff}, denoted as $\hat{d}_{eff}$, can be found as the largest integer such that $\hat{\lambda}_{\hat{d}_{eff}} > \rho_0$. Therefore, the first $\hat{d}_{eff}$ columns of $\hat{T}_y$ give the estimate of V, denoted by $\hat{V}$, and the last $M_t - \hat{d}_{eff}$ columns of $\hat{T}_y$ are deemed as the estimate of U, denoted by $\hat{U}$. Note that $\hat{U}$ will replace the true value of U in (8) for the proposed CB design in the case with imperfect channel learning.

2) Unknown noise power ρ_0: In this case, ρ_0 is unknown to CR-Tx and has to be estimated along with $\hat{Q}_s$. The ML estimate of ρ_0 can first be obtained as [19]

$$\hat{\rho}_0 = \frac{1}{M_t - \hat{d}_{eff}} \sum_{i=\hat{d}_{eff}+1}^{M_t} \hat{\lambda}_i \qquad (11)$$

where $\hat{d}_{eff}$ is the ML estimate of d_{eff}. Specifically, $\hat{d}_{eff}$ can be obtained as [19]

$$\hat{d}_{eff} = \arg\max_k (M_t - k)N \log\left(\frac{\prod_{i=k+1}^{M_t} \hat{\lambda}_i^{1/(M_t - k)}}{\frac{1}{M_t - k}\sum_{i=k+1}^{M_t} \hat{\lambda}_i}\right)$$
$$= \arg\max_k (M_t - k)N \log\left(\frac{\mathrm{GM}(k)}{\mathrm{AM}(k)}\right) \qquad (12)$$

where $\mathrm{GM}(k)$ and $\mathrm{AM}(k)$ denote the geometric mean and the arithmetic mean of the last $M_t - k$ eigenvalues of $\hat{Q}_y$, respectively. To make this estimation unbiased, we conventionally adopt the so-called minimum description length (MDL) estimator expressed as [19]

$$\hat{d}_{eff} = \arg\min_k (M_t - k)N \log\left(\frac{\mathrm{AM}(k)}{\mathrm{GM}(k)}\right) + \frac{1}{2}k(2M_t - k)\log N \qquad (13)$$

where the second term on the right-hand side (RHS) is a bias correction term. The ML estimates of V and U, denoted by

$\hat{V}$ and $\hat{U}$, are then obtained from the first $\hat{d}_{\text{eff}}$ and the last $M_t - \hat{d}_{\text{eff}}$ columns of $\hat{T}_y$, respectively.

After knowing $\hat{\rho}_0$, $\hat{d}_{\text{eff}}$, $\hat{V}$, and $\hat{U}$, the ML estimate of Q_s is obtained as

$$\hat{Q}_s = \hat{V}\text{Diag}\left(\hat{\lambda}_1 - \hat{\rho}_0, \ldots, \hat{\lambda}_{\hat{d}_{\text{eff}}} - \hat{\rho}_0\right)\hat{V}^H. \quad (14)$$

From (10) and (14), it is observed that the above two estimators have a similar structure while they differ in the noise power term adopted and the way to estimate the rank of Q_s, d_{eff}.

B. Effective Leakage Interference

Due to imperfect channel estimation, the CB design in (8) with U replaced by $\hat{U}$ cannot perfectly remove the effective interference at PR_j's. In this subsection, the effect of channel estimation errors on the resultant leakage interference power levels at PR_j's will be analytically quantified so as to assist the later studies. Define the rank over-estimation probability as $p_o(k) = \text{Prob}(\hat{d}_{\text{eff}} - d_{\text{eff}} = k|\hat{d}_{\text{eff}})$, $k = 1, \ldots, \hat{d}_{\text{eff}}$, and the rank under-estimation probability as $p_u(k) = \text{Prob}(d_{\text{eff}} - \hat{d}_{\text{eff}} = k|\hat{d}_{\text{eff}})$, $k = 1, \ldots, M_t - \hat{d}_{\text{eff}}$, both conditioned on the observation $\hat{d}_{\text{eff}}$. If the over-estimation of d_{eff} is encountered, the upper bound on the number of data streams from CR-Tx, d_{CR}, may be affected. However, as long as $(M_t - \hat{d}_{\text{eff}}) \geq M_r$, d_{CR} is more tightly bounded by M_r and the over-estimation of d_{eff} does not cause any problem. On the other hand, the under-estimation of d_{eff} will bring a severe issue, since some columns in $\hat{U}$ may actually come from the PR signal subspace spanned by V. In this case, the amount of interference at PRs will be tremendously increased, which is similar to the scenario in the conventional interleave-based CR system when a misdetection of active PR transmissions occurs to the CR. In practice, a threshold ξ should be properly set, and the last $M_t - (\hat{d}_{\text{eff}} + k_0)$ columns in $\hat{T}_y$ are chosen as $\hat{U}$ only if $p_o(k_0) \geq \xi$.

Detailed study on $p_o(k)$, $p_u(k)$, and ξ is deemed as a separate topic of this paper and will not be further addressed here. In this paper, for simplicity we will assume that the rank of Q_s or d_{eff} is correctly estimated. We will then focus on studying the effect of finite N on the distortion of the estimated eigenspace $\hat{U}$. From (8), the transmit signal at CR-Tx in the case of imperfect channel learning is expressed as

$$s_{\text{CR}}(n) = A_{\text{CR}}t_{\text{CR}}(n) = \hat{U}C_{\text{CR}}^{1/2}t_{\text{CR}}(n), \quad n > N \quad (15)$$

where $s_{\text{CR}}(n)$ is the precoded version of the data vector $t_{\text{CR}}(n)$. Note that $\mathbb{E}[t_{\text{CR}}(n)(t_{\text{CR}}(n))^H] = I$ and $\mathbb{E}[s_{\text{CR}}(n)(s_{\text{CR}}(n))^H] = S_{\text{CR}}$. The average leakage interference power at PR_j, $j = 1, 2$, due to the CR transmission is then expressed as

$$I_j = \mathbb{E}[\|B_j G_j s_{CR}(n)\|^2]. \quad (16)$$

Next, I_j is normalized by the corresponding processed (via B_j) noise power to unify the discussions for PR_j's. For convenience, it is assumed that the additive noise power at PR_j is equal to ρ_0, the same as that at CR-Tx, and thus the processed noise power becomes $\rho_0\text{Tr}(B_j B_j^H)$. Define

$$\bar{I}_j \triangleq \frac{I_j}{\rho_0\text{Tr}(B_j B_j^H)}. \quad (17)$$

$\bar{I}_j$ is then named as the "effective leakage interference power" at PR_j since it measures the power of interference normalized by that of noise after they are both processed by the receive beamforming matrix, B_j.

Lemma 5.1: The upper bounds on $\bar{I}_j$, $j = 1, 2$, are given as

$$\bar{I}_j \leq \frac{\text{Tr}(C_{CR})}{\alpha_j N} \frac{\lambda_{\max}(G_j G_j^H)}{\lambda_{\min}(A_j^H G_j G_j^H A_j)}. \quad (18)$$

Proof: Please refer to Appendix I. ∎

From Lemma 5.1, it follows that the upper bound on $\bar{I}_j$ is proportional to the CR transmit power $\text{Tr}(C_{CR})$, but inversely proportional to α_j, N, and the PR_j's average transmit power P_j (via A_j). Some nice properties of the derived effective leakage interference powers associated with the proposed CB scheme are listed as follows:

- $\bar{I}_j$ is upper-bounded by a finite value provided that $\alpha_j > 0$.[6] Note that $\lambda_{\min}(A_j^H G_j G_j^H A_j) > 0$ if $M_t > d_j$ and in this case $A_j^H G_j$ is a full-rank and fat matrix.
- $\bar{I}_j$ can be easily shown to be invariant to any scalar multiplication over G_j. Thus, the CR protects PR_j's equally regardless of their location-dependent signal attenuation.
- Since for fixed N and P_{CR} the upper bound on $\bar{I}_j$ is inversely proportional to α_j and P_j, PR_j gets better protected if it transmits more frequently and/or with more power. This property is useful for the CR to design *fair* rules for distributing its leakage interference among the coexisting PRs.

Example 5.1: In Figs. 4 (a) and 4 (b), numerical results on $\bar{I}_j$'s given in (17) as well as theoretical results on the upper bounds on $\bar{I}_j$'s given in (18) are compared for PR SNR being 15dB and 0dB, respectively. Note that $P_1 = P_2 = P$ in this example and PR SNR is defined as P/ρ_0, where ρ_0 is assumed to be known at CR-Tx. For the PR, it is assumed that $M_1 = M_2 = 1$, $\alpha_1 = 0.3$, and $\alpha_2 = 0.6$, while for the CR, $M_t = 4$, $P_{\text{CR}} = 100$, and C_{CR} is designed for the eigenmode transmission. $2,000$ random channel realizations are used for averaging while the standard Rayleigh fading channel distribution is adopted. To clearly see the effect of N, we take the inverses of $\bar{I}_j$'s or their upper bounds for the vertical axis of each figure. It is observed that at high-SNR region, the theoretical and numerical results match well, and the interference powers are inversely linearly proportional to N. However, at low-SNR region, there exists big mismatch between the two results. This is reasonable since the first order approximation of (30) in Appendix I is inaccurate at low-SNR region. Nonetheless, the good news is that the inverse of interference power is observed to be still linearly proportional to N from the numerical results.

---

[6]Note that the derived upper bound on $\bar{I}_j$ is practically meaningful when α_j is a non-negligible positive number, since in the extreme case of $\alpha_j = 0$, PR_j switches off its transmissions and as a result discussion for the interference at PR_j becomes irrelevant.
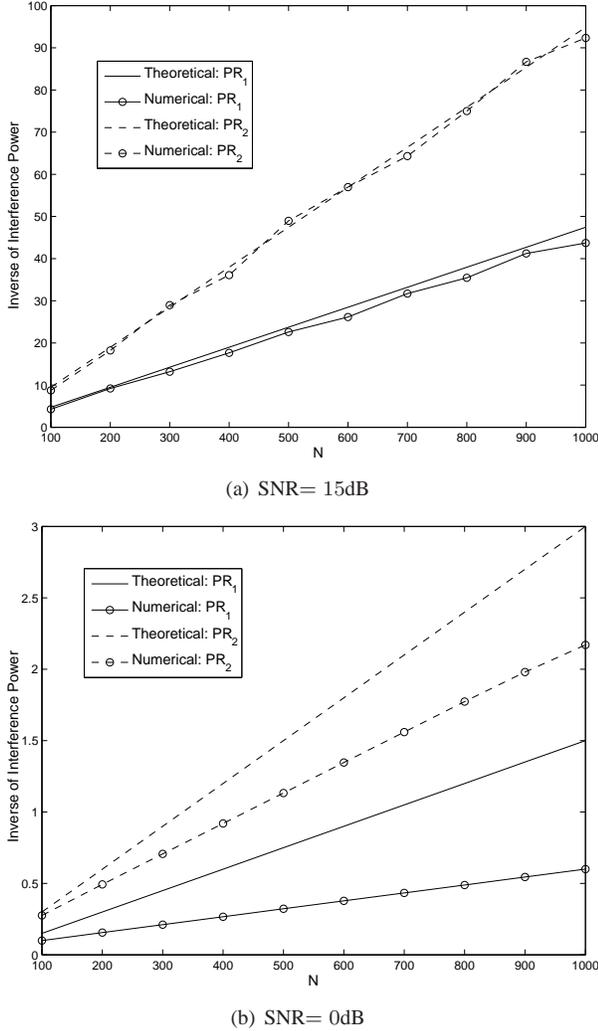
(a) SNR= 15dB

(b) SNR= 0dB

Fig. 4. Leakage interference power levels at PR_1 and PR_2 for different PR SNRs.

C. Optimal Learning Time

At last, we study the learning-throughput tradeoff for CRs by determining the optimal learning time τ for a given T to maximize the CR link throughput, subject to both transmit power constraint of the CR and effective leakage interference power constraints at the two PR terminals. It is assumed that the CR channel $\boldsymbol{H}$ is known at both CR-Tx and CR-Rx. From (8) with $\boldsymbol{U}$ replaced by $\hat{\boldsymbol{U}}$, the maximum CR link effective throughput, assuming the receiver noise $\sim \mathcal{CN}(\boldsymbol{0}, \rho_1 \boldsymbol{I})$, is expressed as [22]

$$\frac{T-\tau}{T} \log \left| \boldsymbol{I} + \boldsymbol{H}\hat{\boldsymbol{U}}\boldsymbol{C}_{CR}\hat{\boldsymbol{U}}^H \boldsymbol{H}^H / \rho_1 \right| \qquad (19)$$

where the term $(T-\tau)/T$ accounts for the portion of throughput loss due to channel learning.

If the *peak* transmit power constraint for the CR is adopted, we have $\text{Tr}(\boldsymbol{C}_{CR}) \leq P_{CR}$, while if the *average* transmit power constraint is adopted, we may allocate the total power for each block to the second phase transmission, resulting in $\text{Tr}(\boldsymbol{C}_{CR}) \leq \frac{T}{T-\tau} P_{CR}$. Let Γ denote the prescribed constraint on the maximum effective leakage interference powers, $\bar{I}_j$'s

defined in (17). Note that N is related to τ by $N = \tau/T_s$, where T_s denotes the symbol period. From Lemma 5.1, it follows that it is sufficient for $\boldsymbol{C}_{CR}$ to satisfy the following inequality to ensure the given interference power constraints:

$$\text{Tr}(\boldsymbol{C}_{CR}) \leq \gamma_j \tau, \ j = 1, 2 \qquad (20)$$

where

$$\gamma_j = \frac{\zeta_j \alpha_j \Gamma}{T_s} \frac{\lambda_{\min}(\boldsymbol{A}_j^H \boldsymbol{G}_j \boldsymbol{G}_j^H \boldsymbol{A}_j)}{\lambda_{\max}(\boldsymbol{G}_j \boldsymbol{G}_j^H)} \qquad (21)$$

and ζ_j, $\zeta_j \leq 1$, is an additional margin that accounts for any analytical errors (e.g., approximations made at low-SNR region in Example 5.1). In practice, the choice of γ_j's in (21) depends on the calibration process at CR-Tx, based on prior knowledge of ζ_j's, Γ, and T_s, as well as the observed average signal power from PRs.

Let $\gamma = \min(\gamma_1, \gamma_2)$. Then, the interference power constraints in (20) become equivalent to $\text{Tr}(\boldsymbol{C}_{CR}) \leq \gamma \tau$. The problem for maximizing the CR effective throughput is thus expressed as

$$(P1): \max_{\tau, \boldsymbol{C}_{CR}} \quad \frac{T-\tau}{T} \log \left| \boldsymbol{I} + \boldsymbol{H}\hat{\boldsymbol{U}}\boldsymbol{C}_{CR}\hat{\boldsymbol{U}}^H \boldsymbol{H}^H / \rho_1 \right|$$
$$\text{s.t.} \quad \text{Tr}(\boldsymbol{C}_{CR}) \leq J, \quad \boldsymbol{C}_{CR} \succeq \boldsymbol{0}, \quad 0 \leq \tau < T$$

where $J = \min(P_{CR}, \gamma\tau)$ for the case of peak transmit power constraint, while $J = \min\left(\frac{T}{T-\tau}P_{CR}, \gamma\tau\right)$ for the case of average transmit power constraint.

For problem (P1), it is noted that $\hat{\boldsymbol{U}}$ is related to τ, which makes the maximization over τ difficult. However, it can be verified that the matrix norm of $\Delta\boldsymbol{U} = \hat{\boldsymbol{U}} - \boldsymbol{U}$ decreases in the order of $\mathcal{O}(1/\sqrt{\tau})$, as compared to the norm of $\boldsymbol{U}$. Therefore, the overall term $\hat{\boldsymbol{U}} = \boldsymbol{U} + \Delta\boldsymbol{U}$ in the objective function is dominated by $\boldsymbol{U}$, and changes slowly with τ when τ is sufficiently large. Thus, we assume that the effect of τ on $\hat{\boldsymbol{U}}$ is ignored in the subsequent analysis, and will verify this assumption via simulation results in Section VI.

Let the EVD of $\hat{\boldsymbol{U}}^H \boldsymbol{H}^H \boldsymbol{H}\hat{\boldsymbol{U}}$ be $\boldsymbol{U}_h \boldsymbol{\Sigma}_h \boldsymbol{U}_h^H$, where $\boldsymbol{U}_h$ is a $(M_t - d_{\text{eff}}) \times (M_t - d_{\text{eff}})$ unitary matrix and $\boldsymbol{\Sigma}_h = \text{Diag}(\sigma_{h,1}^2, \ldots, \sigma_{h,M_t-d_{\text{eff}}}^2)$. W.l.o.g., we assume that $\sigma_{h,i}^2$'s are arranged in a descending order. Note that if $(M_t - d_{\text{eff}}) > M_r$, then $\sigma_{h,i}$'s, $i = M_r + 1, \ldots, M_t - d_{\text{eff}}$, all have zero values. Define $\boldsymbol{X}$ as $\boldsymbol{U}_h^H \boldsymbol{C}_{\text{CR}} \boldsymbol{U}_h$. Problem (P1) is then converted to

$$(P2): \max_{\tau, \boldsymbol{X}} \quad \frac{T-\tau}{T} \log |\boldsymbol{I} + \boldsymbol{X}\boldsymbol{\Sigma}_h / \rho_1|$$
$$\text{s.t.} \quad \text{Tr}(\boldsymbol{X}) \leq J, \quad \boldsymbol{X} \succeq \boldsymbol{0}, \quad 0 \leq \tau < T$$

where the optimal $\boldsymbol{C}_{\text{CR}}$ can be later recovered as $\boldsymbol{U}_h \boldsymbol{X} \boldsymbol{U}_h^H$. By the standard approach like in [22, Chapter 10.5], it can be shown that the optimal $\boldsymbol{X}$ is a diagonal matrix $\boldsymbol{X} = \text{Diag}(x_1, \ldots, x_{M_t-d_{\text{eff}}})$ and x_i's, $i = 1, \ldots, M_t - d_{\text{eff}}$, are obtained from

$$(P3): \max_{\tau, \{x_i\}} \quad \frac{T-\tau}{T} \sum_{i=1}^{M_t-d_{\text{eff}}} \log \left(1 + \frac{\sigma_{h,i}^2 x_i}{\rho_1} \right)$$
$$\text{s.t.} \quad \sum_{i=1}^{M_t-d_{\text{eff}}} x_i \leq J, \quad x_i \geq 0, \quad 0 \leq \tau < T.$$

Next, we will study (P3) for the cases of peak and average transmit power constraints, respectively.

1) Peak transmit power constraint: In this case, if $P_{CR} > \gamma T$, then J is always equal to $\gamma\tau$. Therefore, we consider the more general case with $P_{CR} \le \gamma T$. The remaining discussion will then be divided into the following two parts for $P_{CR}/\gamma < \tau < T$ and $0 \le \tau \le P_{CR}/\gamma$, respectively.

If $P_{CR}/\gamma < \tau < T$, then $J = P_{CR}$ and the optimization in problem (P3) over τ and x_i's can be separated. The optimization over x_i's directly follows the conventional water-filling (WF) algorithm [22]. For the ease of later discussion, we define

$$f(z) = \max_{\{x_i\}} \sum_{i=1}^{M_t - d_{\text{eff}}} \log\left(1 + \frac{\sigma_{h,i}^2 x_i}{\rho_1}\right)$$
$$\text{s.t.} \sum_{i=1}^{M_t - d_{\text{eff}}} x_i \le z, \quad x_i \ge 0. \tag{22}$$

The WF solution of the above optimization problem is then given as $x_i = (\frac{1}{\mu} - \frac{\rho_1}{\sigma_{h,i}^2})^+$, where $\frac{1}{\mu}$ is the water level that should satisfy

$$\sum_{i=1}^{M_t - d_{\text{eff}}} \left(\frac{1}{\mu} - \frac{\rho_1}{\sigma_{h,i}^2}\right)^+ = z. \tag{23}$$

Denote $q_k = \frac{k\rho_1}{\sigma_{h,k+1}^2} - \sum_{i=1}^{k} \frac{\rho_1}{\sigma_{h,i}^2}$, for $k = 0, \ldots, M_t - d_{\text{eff}}$. Obviously, $q_0 = 0$, and $q_{M_t - d_{\text{eff}}} = +\infty$ since $\sigma_{h,M_t - d_{\text{eff}}+1}^2$ is set to be zero. Then, we can express $f(z)$ as

$$f(z) = \sum_{i=1}^{k} \log\left(\frac{\sigma_{h,i}^2}{k\rho_1}\left(z + \sum_{i=1}^{k} \frac{\rho_1}{\sigma_{h,i}^2}\right)\right), \quad z \in [q_{k-1}, q_k]. \tag{24}$$

Note that k is the number of dimensions assigned with positive x_i's. The objective function of problem (P3) in this case can then be explicitly written as

$$g_1(\tau) \triangleq \frac{T-\tau}{T} f(P_{CR}). \tag{25}$$

Since $\frac{T-\tau}{T}$ is a decreasing function of τ, the optimal τ to maximize $g_1(\tau)$ over $P_{CR}/\gamma < \tau \le T$ is simply P_{CR}/γ.

Next, consider $0 \le \tau \le P_{CR}/\gamma$. In this case, $J = \gamma\tau$, and problem (P3) becomes

$$\max_{0 \le \tau \le P_{CR}/\gamma} g_2(\tau) \triangleq \frac{T-\tau}{T} f(\gamma\tau). \tag{26}$$

In order to study the function $g_2(\tau)$, some properties of the function $f(z)$ are given below.

Lemma 5.2: $f(z)$ is a continuous, increasing, differentiable, and concave function of z.

Proof: Please refer to Appendix II. ∎

With Lemma 5.2, it can be easily verified that $g_2(\tau)$ is also a continuous, differentiable, and concave function of τ. Thus, the optimal value of τ, denoted as τ_2^*, to maximize $g_2(\tau)$ can be easily obtained via, e.g., the Newton method [23].

To summarize the above two cases, the optimal solution of τ for problem (P3) in the case of peak transmit power constraint can be obtained as

$$\tau^* = \begin{cases} \tau_2^*, & \tau_2^* < P_{CR}/\gamma \\ P_{CR}/\gamma, & \text{otherwise.} \end{cases} \tag{27}$$
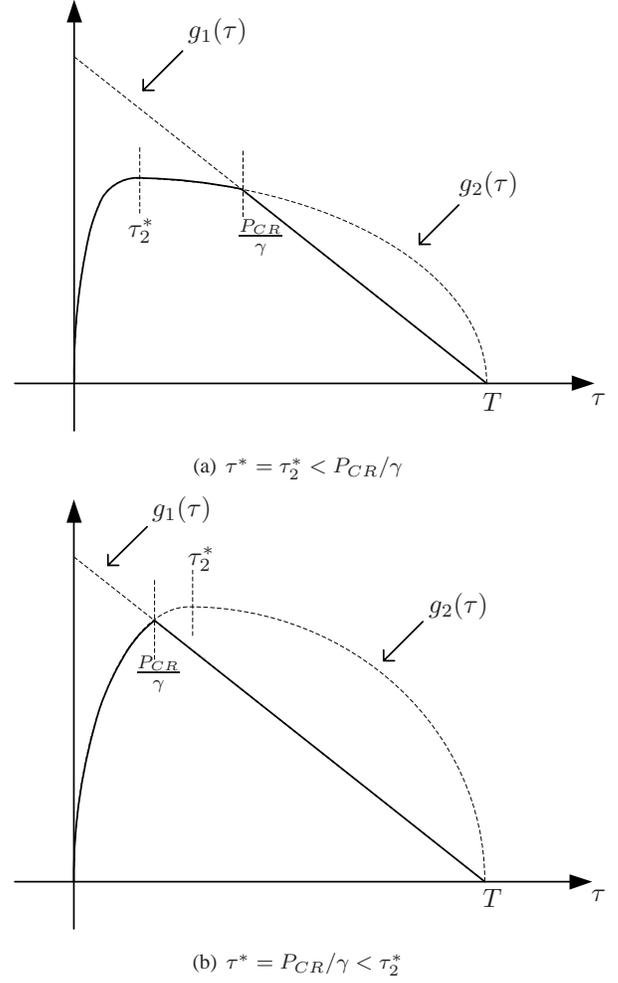

(a) $\tau^* = \tau_2^* < P_{CR}/\gamma$

(b) $\tau^* = P_{CR}/\gamma < \tau_2^*$

Fig. 5. Illustration of the optimal learning time τ^* for the case of peak CR transmit-power constraint.

The above solution is illustrated in Fig. 5. The optimal value of (P3) then becomes $g_2(\tau_2^*)$ if $\tau_2^* < P_{CR}/\gamma$, and $g_1(P_{CR}/\gamma)$ otherwise.

2) Average transmit power constraint: In this case, J in problem (P3) takes the value of $T/(T-\tau)P_{CR}$ if $T/(T-\tau)P_{CR} < \gamma\tau$, and $\gamma\tau$ otherwise. It can be verified that $T/(T-\tau)P_{CR} < \gamma\tau$ for some τ in $[0, T)$ only if $P_{CR}/\gamma < T/4$. In other words, if $P_{CR}/\gamma \ge T/4$, J always takes the value $\gamma\tau$ regardless of τ. Thus, the objective function of (P3) is always given as $g_2(\tau)$, and the optimal solution of τ is τ_2^*.

Therefore, we consider next the more general case of $P_{CR}/\gamma < T/4$. For this case, it can be shown that the equation $T/(T-\tau)P_{CR} = \gamma\tau$ always has two positive roots of τ, denoted as τ_l and τ_u, respectively, and $0 \le \tau_l < \tau_u < T$. If $0 \le \tau \le \tau_l$ or $\tau_u \le \tau < T$, J takes the value of $\gamma\tau$, and then the maximum value of (P3) is obtained by the τ that maximizes $g_2(\tau)$ over this interval of τ. Otherwise, the maximum value occurs when τ is given as

$$\arg\max_{\tau, \tau_l < \tau < \tau_u} g_3(\tau) \triangleq \frac{T-\tau}{T} f\left(\frac{T}{T-\tau} P_{CR}\right). \tag{28}$$

It can be shown that $g_3(\tau)$ is a continuously decreasing function of τ, for $\tau \in [0, T)$. Thus, the optimal value of τ
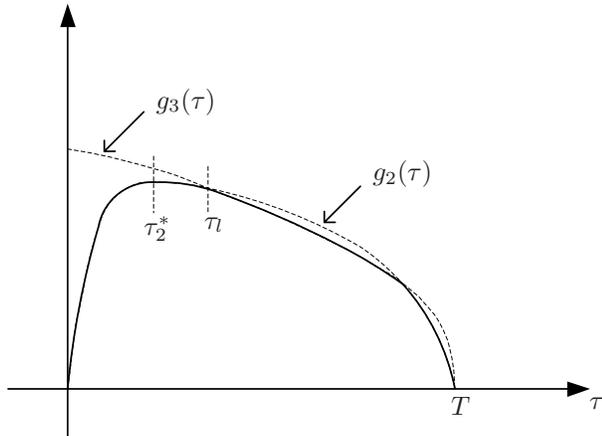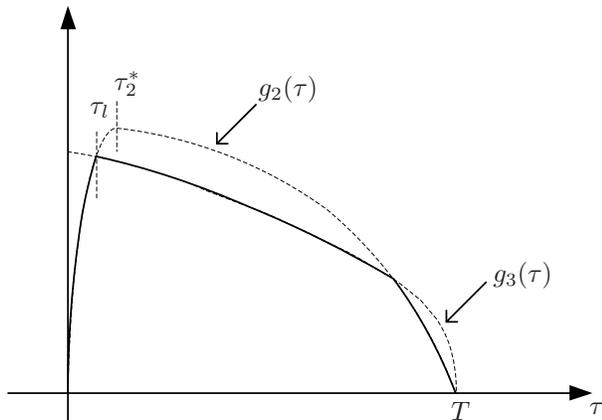
(a) $\tau^* = \tau_2^* < \tau_l$

(b) $\tau^* = \tau_l < \tau_2^*$

Fig. 6. Illustration of the optimal learning time τ^* for the case of average CR transmit-power constraint.

to maximize $g_3(\tau)$ over this interval of τ is simply τ_l.

To summarize the above discussions, we obtain the optimal solution of τ for problem (P3) in the case of average transmit power constraint as

$$\tau^* = \begin{cases} \tau_2^*, & \tau_2^* < \tau_l \\ \tau_l, & \text{otherwise.} \end{cases} \quad (29)$$

The above solution is illustrated in Fig. 6. The optimal value of (P3) then becomes $g_2(\tau_2^*)$ if $\tau_2^* < \tau_l$, and $g_3(\tau_l)$ otherwise.

VI. NUMERICAL RESULTS

In the section, we present additional simulation results to demonstrate the performance of the proposed CB scheme under imperfect channel learning. The system parameters are taken as $M_t = 6$, $M_r = 3$, $M_1 = 4$, and $M_2 = 2$. Eigenmode transmission is considered for the PR with $d_1 = d_2 = 2$, and the PR SNR is set as 20dB. The channels $\boldsymbol{F}$, $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, and $\boldsymbol{H}$ are randomly generated from the standard Rayleigh fading distribution, and are then fixed in all the examples. The parameters τ and T are normalized by the symbol period T_s. After the normalization, T is set as $1,000$ and the lowest value of τ is set as 10. The CR transmit rate is measured
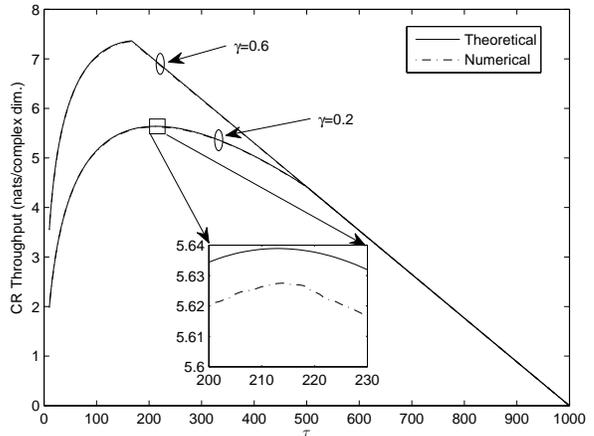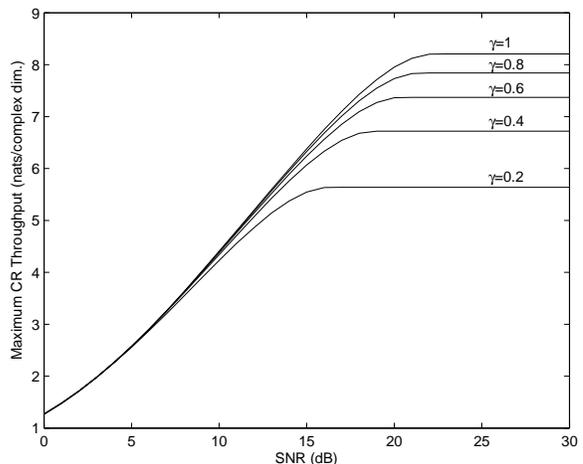

Fig. 7. CR throughput versus CR learning time.

Fig. 8. Maximum CR throughput versus CR SNR.

in nats/complex dimension (dim.). The peak transmit-power constraint for the CR is assumed.

We first fix P_{CR} at CR-Tx as 100 and show the variations of the CR throughput as a function of τ. Both theoretical results obtained in Section V-C where $\hat{\boldsymbol{U}}$ is not considered as a function of τ and is replaced by the true value $\boldsymbol{U}$, and numerical results where $\hat{\boldsymbol{U}}$ is obtained via the estimator given in Section V-A with known noise power ρ_0, are shown in Fig. 7. The values of γ are taken as 0.2 and 0.6, respectively. From Fig. 7, the first observation is that the numerical and theoretical results almost merge with each other, which supports our previous assumption of ignoring $\hat{\boldsymbol{U}}$ to be a function of τ during the optimization process. We also observe that the CR throughput for $\gamma = 0.2$ and that for $\gamma = 0.6$ start to merge when τ is sufficiently large due to the fact that $g_1(\tau)$ defined in (25) does not change with γ. However, the maximum CR throughput is observed to increase with γ because when the PRs can tolerate more leakage interference powers, the optimal learning time is reduced and the CR transmit power becomes less restricted, which leads to an increased CR throughput.
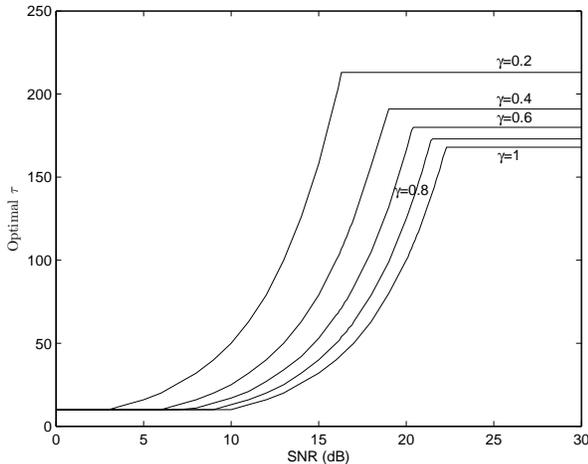
Fig. 9.   Optimal CR learning time versus CR SNR.

We then display the maximum CR throughput versus P_{CR}, or equivalently, the CR SNR, in Fig. 8 for different values of γ. Only the theoretical results are shown here. The first observation is that there exist thresholds on CR SNR, beyond which the maximum CR throughput cannot be improved for a given γ. This is because that when P_{CR} is too large, the dominant constraint for CR throughput maximization becomes the interference-power constraint instead of transmit-power constraint. When this occurs, the intersection point P_{CR}/γ in Fig. 5 moves towards T. Thus, the optimal value of τ and the corresponding maximum CR throughput are determined from $g_2(\tau)$ in (26), which is not related to P_{CR}. Meanwhile, when γ increases, the maximum CR throughput also increases, similarly like the case in Fig. 7.

At last, we show the change of the optimal τ with respect to P_{CR} or the CR SNR in Fig. 9, where only the theoretical results are shown. From Fig. 5, we know that when P_{CR} decreases, the intersection point moves towards zero. Thus, the curves of the optimal learning time for different γ's all merge to the presumed minimum value for τ, $\tau = 10$, at low-SNR region. On the other side, the optimal values of τ stop increasing at high-SNR region for a given γ, similarly as explained for Fig. 8. Moreover, the optimal τ is observed to increase with the decreasing of γ.

VII. CONCLUDING REMARKS

Cognitive beamforming (CB) is a promising technology to enable high-rate CR transmissions and yet to provide effective interference avoidance at the coexisting PR terminals. The main challenge for realizing CB in practical systems is how to obtain the channel knowledge from CR transmitter to PR terminals. In this paper, we propose a new solution to this problem utilizing the idea of effective interference channel (EIC), which can be efficiently learned at CR transmitter via blind/semiblind estimation over the received PR signals. Based on the EIC, we design a practical CB scheme to minimize the effect of the resulted interference on the PR transmissions. Furthermore, we show that with finite sample size for channel

learning, there exists an optimal learning time to maximize the CR link throughput.

The developed results in this paper can be readily extended to the case with multiple PR links. This is so because the proposed CB scheme is based on the EIC that measures the space spanned by all the coexisting PR signals as a whole, and thus it works regardless of these PR signals coming from a single PR link or multiple PR links.

APPENDIX I
PROOF OF LEMMA 5.1

Define $\boldsymbol{S} = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(N)]$ and $\boldsymbol{Y}_s = \mathcal{A}\boldsymbol{S}$ where $\boldsymbol{s}(n)$'s and $\mathcal{A}$ are given in Section III. From [20, Appendix I], we know that the first order perturbation[7] to $\boldsymbol{U}$ due to the finite number of samples N and the additive noise $\boldsymbol{Z} \triangleq [\boldsymbol{z}(1), \ldots, \boldsymbol{z}(N)]$ can be approximated by

$$\Delta \boldsymbol{U} \triangleq \hat{\boldsymbol{U}} - \boldsymbol{U} \approx -(\boldsymbol{Y}_s^H)^\dagger \boldsymbol{Z}^H \boldsymbol{U}. \tag{30}$$

Since the discussions on $\bar{I}_1$ and $\bar{I}_2$ are similar, in the following we restrict our study on $\bar{I}_1$. From the conditions given in Proposition 4.1, we know that there exists a constant matrix $\boldsymbol{W}_1 \in \mathbb{C}^{d_2 \times d_1}$, such that $\boldsymbol{B}_1 \boldsymbol{G}_1 = \boldsymbol{W}_1 \boldsymbol{A}_1^H \boldsymbol{G}_1$. The average interference power, I_1 defined in (16), is then re-expressed as

$$
\begin{aligned}
I_1 &\overset{(a)}{=} \mathbb{E}[\text{Tr}(\boldsymbol{B}_1 \boldsymbol{G}_1 \hat{\boldsymbol{U}} \boldsymbol{C}_{CR} \hat{\boldsymbol{U}}^H \boldsymbol{G}_1^H \boldsymbol{B}_1^H)] \\
&\overset{(b)}{=} \mathbb{E}[\text{Tr}(\boldsymbol{B}_1 \boldsymbol{G}_1 \Delta \boldsymbol{U} \boldsymbol{C}_{CR} \Delta \boldsymbol{U}^H \boldsymbol{G}_1^H \boldsymbol{B}_1^H)] \\
&\overset{(c)}{=} \mathbb{E}[\text{Tr}(\boldsymbol{B}_1 \boldsymbol{G}_1 (\boldsymbol{Y}_s^H)^\dagger \boldsymbol{Z}^H \boldsymbol{U} \boldsymbol{C}_{CR} \boldsymbol{U}^H \boldsymbol{Z} \boldsymbol{Y}_s^\dagger \boldsymbol{G}_1^H \boldsymbol{B}_1^H)] \\
&\overset{(d)}{=} \rho_0 \text{Tr}(\boldsymbol{C}_{CR}) \mathbb{E}[\text{Tr}(\boldsymbol{B}_1 \boldsymbol{G}_1 (\boldsymbol{Y}_s^H)^\dagger \boldsymbol{Y}_s^\dagger \boldsymbol{G}_1^H \boldsymbol{B}_1^H))] \\
&\overset{(e)}{=} \rho_0 \text{Tr}(\boldsymbol{C}_{CR}) \mathbb{E}\Big[\text{Tr}(\boldsymbol{W}_1 \boldsymbol{A}_1^H \boldsymbol{G}_1 (\mathcal{A}^H)^\dagger \times \\
&\qquad (\boldsymbol{S}\boldsymbol{S}^H)^{-1} \mathcal{A}^\dagger \boldsymbol{G}_1^H \boldsymbol{A}_1 \boldsymbol{W}_1^H)\Big] \\
&\overset{(f)}{\approx} \rho_0 \text{Tr}(\boldsymbol{C}_{CR}) \text{Tr}\left(\boldsymbol{W}_1 [\boldsymbol{I}, \boldsymbol{0}] \begin{bmatrix} \frac{1}{|\mathcal{N}_1|}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{|\mathcal{N}_2|}\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{W}_1^H\right) \\
&= \frac{\rho_0}{\alpha_1 N} \text{Tr}(\boldsymbol{C}_{CR}) \text{Tr}(\boldsymbol{W}_1 \boldsymbol{W}_1^H) \tag{31}
\end{aligned}
$$

where (a) is via substituting (15) into (16) and using the independence of $\hat{\boldsymbol{U}}$ and $\boldsymbol{t}_{\text{CR}}(n)$; (b) is due to $\boldsymbol{B}_1 \boldsymbol{G}_1 \boldsymbol{U} = \boldsymbol{0}$; (c) is due to (30); (d) is due to independence of $\boldsymbol{Y}_s$ and $\boldsymbol{Z}$ and $\mathbb{E}[\boldsymbol{Z}^H \boldsymbol{X} \boldsymbol{Z}] = \rho_0 \text{Tr}(\boldsymbol{X}) \boldsymbol{I}$ for any constant matrix $\boldsymbol{X}$; (e) is due to the definitions of $\boldsymbol{W}_1$ and $\boldsymbol{Y}_s$; and (f) is approximately true since N is usually a large number.

From [21], we have

$$
\begin{aligned}
&\text{Tr}(\boldsymbol{W}_1 \boldsymbol{A}_1^H \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{A}_1 \boldsymbol{W}_1^H) \\
&\geq \lambda_{\min}(\boldsymbol{A}_1^H \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{A}_1) \text{Tr}(\boldsymbol{W}_1 \boldsymbol{W}_1^H) \tag{32}
\end{aligned}
$$

$$\text{Tr}(\boldsymbol{B}_1 \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{B}_1^H) \leq \lambda_{\max}(\boldsymbol{G}_1 \boldsymbol{G}_1^H) \text{Tr}(\boldsymbol{B}_1 \boldsymbol{B}_1^H). \tag{33}$$

By noting $\boldsymbol{B}_1 \boldsymbol{G}_1 = \boldsymbol{W}_1 \boldsymbol{A}_1^H \boldsymbol{G}_1$, from (32) and (33) it follows that

$$\text{Tr}(\boldsymbol{W}_1 \boldsymbol{W}_1^H) \leq \frac{\lambda_{\max}(\boldsymbol{G}_1 \boldsymbol{G}_1^H) \text{Tr}(\boldsymbol{B}_1 \boldsymbol{B}_1^H)}{\lambda_{\min}(\boldsymbol{A}_1^H \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{A}_1)}. \tag{34}$$

---

[7]Note that the first order approximation is more valid at high-SNR region.

Using (17), (31), and (34), the upper bound on $\bar{I}_1$ given in (18) is obtained.

APPENDIX II
PROOF OF LEMMA 5.2

First, it is easily known that $f(z)$ is an increasing function of z. Next, we prove the continuity, differentiability, and concavity of $f(z)$, respectively.

A. Continuity

From (24), it is known that in each section $[q_{k-1}, q_k]$, $f(z)$ is obviously continuous. For boundary points of each section, we have

$$\lim_{z \to q_k^-} f(z) = \sum_{i=1}^{k} \log\left(\frac{\sigma_{h,i}^2}{\sigma_{h,k+1}^2}\right) = \lim_{z \to q_k^+} f(z), \quad (35)$$

$k = 1, \ldots, M_t - d_{\text{eff}} - 1$. Thus, $f(z)$ is continuous at all the points.

B. Differentiability

From (24), it is known that in each section $[q_{k-1}, q_k]$, $f(z)$ is differentiable. For boundary points of each section, it can be verified that

$$\lim_{z \to q_k^-} \dot{f}(z) = \frac{\sigma_{h,k+1}^2}{\rho_1} = \lim_{z \to q_k^+} \dot{f}(z), \quad (36)$$

$k = 1, \ldots, M_t - d_{\text{eff}} - 1$. Therefore, $f(z)$ is differentiable at all the points.

C. Concavity

For a given z, $f(z)$ is obtained by solving the optimization problem in (22), which can be easily verified to be a convex optimization problem [23]. Thus, the duality gap for this optimization problem is zero and $f(z)$ can be equivalently obtained as the optimal value of the following min-max optimization problem:

$$f(z) = \min_{\mu \geq 0} \max_{x_i \geq 0} \sum_i \log(1 + \frac{\sigma_{h,i}^2 x_i}{\rho_1}) - \mu(\sum_i x_i - z) \quad (37)$$

$$= \min_{\mu \geq 0} \sum_i (\log(\frac{\sigma_{h,i}^2}{\rho_1 \mu}))^+ - \sum_i (1 - \frac{\rho_1 \mu}{\sigma_{h,i}^2})^+ + \mu z \quad (38)$$

$$= \sum_i (\log(\frac{\sigma_{h,i}^2}{\rho_1 \mu^{(z)}}))^+ - \sum_i (1 - \frac{\rho_1 \mu^{(z)}}{\sigma_{h,i}^2})^+ + \mu^{(z)} z \quad (39)$$

where the summations are taken over $i = 1, \ldots, M_t - d_{\text{eff}}$, and $\mu^{(z)} \geq 0$ is the optimal dual variable for a given z. In fact, it can be shown that $1/\mu^{(z)}$ is just the water level given in (23) corresponding to the total power z.

Denote ω as any constant in $[0, 1]$. Let $\mu^{(z_1)}$, $\mu^{(z_2)}$, and $\mu^{(z_3)}$ be the optimal μ for $f(z_1)$, $f(z_2)$, and $f(z_3)$, $z_3 = \omega z_1 + (1 -$

$\omega) z_2$, respectively. For $j = 1, 2$, we have

$$f(z_j) = \sum_i (\log(\frac{\sigma_{h,i}^2}{\rho_1 \mu^{(z_j)}}))^+ - \sum_i (1 - \frac{\rho_1 \mu^{(z_j)}}{\sigma_{h,i}^2})^+ + \mu^{(z_j)} z_j \tag{40}$$

$$\leq \sum_i (\log(\frac{\sigma_{h,i}^2}{\rho_1 \mu^{(z_3)}}))^+ - \sum_i (1 - \frac{\rho_1 \mu^{(z_3)}}{\sigma_{h,i}^2})^+ + \mu^{(z_3)} z_j \tag{41}$$

where the inequality is due to the fact that $\mu^{(z_3)}$ is not the optimal dual solution for $j = 1, 2$. Therefore,

$$\omega f(z_1) + (1 - \omega) f(z_2)$$

$$\leq \sum_i (\log(\frac{\sigma_{h,i}^2}{\rho_1 \mu^{(z_3)}}))^+ - \sum_i (1 - \frac{\rho_1 \mu^{(z_3)}}{\sigma_{h,i}^2})^+ + \mu^{(z_3)} z_3 \quad (42)$$

$$= f(z_3) \tag{43}$$

$$= f(\omega z_1 + (1 - \omega) z_2). \tag{44}$$

Thus, $f(z)$ is a concave function [23].

REFERENCES

[1] J. Mitola, "Cognitive radio: an integrated agent architecture for software defined radio," *PhD Dissertation, KTH, Stockholm, Sweden*, Dec. 2000.

[2] A. Goldsmith, S. A. Jafar, I. Marić, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894-914, May 2009.

[3] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813-1827, May 2006.

[4] A. Jovičić and P. Viswanath, "Cognitive radio: an information-theoretic perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2006.

[5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[6] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 471-487, Feb. 2007.

[7] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 649-658, Feb. 2007.

[8] L. Musavian and S. Aissa, "Capacity and power allocation for spectrum-sharing communications in fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 148-156, Jan. 2009.

[9] X. Kang, Y. C. Liang, A. Nallanathan, H. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: ergodic capacity and outage capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 940-950, Feb. 2009.

[10] R. Zhang, "On peak versus average interference power constraints for protecting primary users in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 2112-2120, Apr. 2009.

[11] Y. Chen, G. Yu, Z. Zhang, H. H. Chen, and P. Qiu, "On cognitive radio networks with opportunistic power control strategies in fading channels," *IEEE. Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2752-2761, Jul. 2008.

[12] R. Zhang, "Optimal power control over fading cognitive radio channels by exploiting primary user CSI," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2008.

[13] R. Zhang, S. Cui, and Y.-C. Liang, "On ergodic sum capacity of fading cognitive multiple-access and broadcast channels," *to appear in IEEE Trans. Inf. Theory*. Available [Online] at arXiv:0806.4468.

[14] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Sig. Process.*, vol. 2, no. 1, pp. 88-102, Feb. 2008.

[15] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[16] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Sig. Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.

[17] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326-1337, Apr. 2008.

[18] T. J. Lim, R. Zhang, Y.-C. Liang, and Y. Zeng, "GLRT-based spectrum sensing for cognitive radio," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2008.

[19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Sig. Process.*, vol. 33, no. 2, pp. 387-392, Apr. 1985.

[20] F. Gao, Y. Zeng, A. Nallanathan, and T.-S. Ng, "Robust subspace blind channel estimation for cyclic prefixed MIMO ODFM systems: algorithm, identifiability and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 378-388, Feb. 2008.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[22] T. Cover and J. Thomas, *Elements of Information Theory,* New York: Wiley, 1991.

[23] S. Boyd and L. Vandenberghe, *Convex optimization,* Cambridge University Press, 2004.