Prescient Beamforming by Primary Transmitters in Interweave Cognitive Radio Networks with a Single Primary Receiver

Amitav Mukherjee, and A. Lee Swindlehurst, *Fellow, IEEE*

Abstract

This work investigates a fundamentally novel interweave cognitive radio network where the primary transmitter takes a proactive approach towards improving the accuracy of the spectrum sensing outcomes at the secondary users (SUs). For the single-primary-receiver scenario considered here, the multi-antenna primary user constructs its transmit beamforming vector so as to increase the detection probability at the SUs while meeting a desired Quality-of-Service (QoS) target on its own link, by exploiting either partial or statistical channel state information of the SUs. The objective of such a proactive approach, which we refer to as *prescient* precoding, is to minimize the probability of interference from SUs at the primary receiver due to imperfect spectrum sensing in fading channels. We also develop information-theoretic bounds on the performance of prescient transmission and study non-linear precoding schemes that approach them. Numerical results are presented to verify the advantages of the proposed prescient transmission techniques for both non-cooperative and cooperative spectrum sensing scenarios.

Index Terms

Cognitive radios, spectrum sensing, interference mitigation, beamforming, dirty-paper coding.

I. INTRODUCTION

Cognitive radios are emerging as promising solutions to enable better utilization of the radio spectrum, especially in bands that are currently under-utilized. The canonical model of a cognitive radio (CR) is one that employs "interweave" cognition [1]. Under the interweave paradigm, cognitive radios seek to opportunistically occupy a channel (frequency band) only when it is not occupied by a primary transmitter (PT) licensed to use that band. In the absence of pre-defined control channels or coordinated medium access between the primary and unlicensed users, the CRs must periodically sense the spectrum for the presence of PTs [2] and cease transmission upon detection. Inevitably, imperfect CR spectrum sensing due to channel fading and other impairments will lead to unintentional interference at the primary receivers (PRs).

The use of multiple antennas in CRs has been suggested for improved spectrum sensing capabilities in interweave systems by means of receive diversity [3]–[5]. MIMO CRs have also been investigated in the context of "underlay" cognitive radio networks, where CRs and PTs coexist in the same spectrum and spectrum sensing is not required. In this case, multiple transmit antennas are used in the secondary network for beamforming and to minimize the interference to the PRs, with complete or partial channel state information (CSI) at the SU transmitter [6]-[10].

In all prior work on interweave networks, the primary transmitters are assumed to be legacy users that are oblivious to the presence of the secondary users. However, the problem considered in this work is fundamentally different in that the PTs construct their transmit beamformers with the objective of increasing the detection probability at the secondary users (SUs) while meeting a desired QoS target on its own link, given partial CSI of the PRs and SUs. Therefore, the primary transmitter adopts a proactive approach, which we refer to as *prescient beamforming*, to minimize the probability of interference from

The authors are with the Dept. of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697-2625, USA. (e-mail: {amukherj; swindle}@uci.edu)

This work was supported by the National Science Foundation under grant CCF-0916073.

SUs at the primary receiver due to imperfect spectrum sensing. Furthermore, we analyze an interweave network where the primary transmitter is equipped with multiple antennas, while the primary receiver(s) and secondary transmitters/receivers have a single antenna each. The potential benefits of multi-antenna PTs and PRs have received surprisingly little attention in the literature on both interweave and underlay CR systems.

The paper is organized as follows. Section II introduces the mathematical model of the network and summarizes the sensing strategy in place at the SUs with emphasis on non-cooperative energy detection, followed by extensions to cooperative spectrum sensing. Section III examines the impact of SU mis-detections on the primary receiver. Section IV presents the prescient beamforming algorithms in detail for a single primary link and multiple SUs. Selected numerical examples are shown in Section V, and we conclude in Section VI.

II. Mathematical Model

A. Signal and Network Model

Consider a network with a single N-antenna PT, its intended PR with a single antenna, and K secondary users (SUs, also synonymous with secondary transmitters), as depicted in Fig. 1. Let the primary terminal be denoted as user 0, and to begin with, assume that we restrict attention to linear transmit beamforming. When active, the PT transmit signal at time index t is given by

$$\mathbf{x}\left(t\right) = \mathbf{w} s_0\left(t\right) \tag{1}$$

where $\mathbf{w} \in \mathbb{C}^{N\times 1}$ is the transmit beamforming weight vector and $s_0\left(t\right)$ is the complex scalar zero-mean information symbol transmitted at t, with power $E\left\{\left|s_0\left(t\right)\right|^2\right\} = \sigma_s^2$. A constraint is assumed to be imposed on the average transmit power of the PT:

$$E\left\{\mathrm{Tr}\left(\mathbf{x}\left(t\right)\mathbf{x}\left(t\right)^H\right)\right\} \leq P \quad \forall t. \tag{2}$$

When the PT is active, the signal received by the k^{th} SU if in sensing mode is

$$y_k\left(t\right) = \mathbf{h}_k\left(t\right)\mathbf{x}\left(t\right) + z_k\left(t\right), \tag{3}$$

where $\mathbf{h}_k\left(t\right)$ is the $1 \times N$ flat-fading complex channel vector between the PT and SU k, and $z_k\left(t\right)$ is a circularly symmetric zero-mean Gaussian noise sample with variance σ_k^2. Multi-antenna SUs can be accommodated in this model assuming they employ a fixed receive beamformer (e.g., a maximal-ratio combiner) prior to the spectrum-sensing stage [4].

We list below the major assumptions regarding the primary and SU network.

- Two different cases regarding the availability of CSI at the PT will be assumed. These are referred to as (1) *partial CSI*, which is defined to mean that the PT has knowledge of the instantaneous realizations of the PT-PR ($\mathbf{h}_0$) and PT-SU ($\{\mathbf{h}_k\}_{k=1}^K$) channels but only the statistics of the SU-PR channels ($\{g_k\}_{k=1}^K$), and (2) *statistical CSI*, which will indicate that the PT only has knowledge of the statistics of all channels in the network.
- The PT is aware of the statistics of the background noise for all users.
- The PT has knowledge of the SU transmit powers and the structure and parameters of the spectrum sensing scheme deployed at the SUs, which in practice are likely to be pre-defined by spectrum regulatory agencies.
- The PT traffic is bursty and its traffic statistics are unknown to the SU spectrum sensors.
- There is no coordination between the PT and SUs in terms of power control.
- The PT terminals and all SUs are half-duplex, which precludes for example simultaneous data transmission and spectrum sensing by the SUs.
- In the non-cooperative sensing scenario the SU network is asynchronous, which implies that the SUs are not all simultaneously in spectrum-sensing mode. We capture this assumption by assigning a

probability β to the event that an arbitrary SU is in spectrum-sensing mode, and a probability $(1-\beta)$ that a SU is engaged in data transmission.

- We only consider in-band spectrum sensing, i.e., sensing is conducted on the same band that is used for data transmission. There is no provision for out-of-band sensing on a separate beacon channel.
- The primary receiver employs single-user decoding and treats all SU interference as noise.
- The interference from the SUs is assumed to be instantaneous, i.e., the processing delay due to spectrum sensing is neglected.

B. Non-cooperative spectrum sensing

In a non-cooperative SU scenario, the SUs attempt to individually determine the presence of an active PT on their communications channel by means of spectrum sensing. The local hypothesis test for spectrum sensing based on $\tilde{M}$ discrete-time samples at the k^{th} secondary user is

$$\begin{array}{lll} \mathcal{H}_0 : & y_k[n] = z_k[n] & n = 0, 1, \ldots, \tilde{M}-1 \\ \mathcal{H}_1 : & y_k[n] = \mathbf{h}_k \mathbf{w} s_0[n] + z_k[n] & n = 0, 1, \ldots, \tilde{M}-1 \end{array} \tag{4}$$

where the channel $\mathbf{h}_k$ is assumed to be constant over the $\tilde{M}$ samples. The $\tilde{M}$ complex samples are composed of $M \triangleq 2\tilde{M}$ independent real and imaginary components [13]. We assume that the background noise at the SUs is temporally uncorrelated. We begin our development by analyzing the detection probability $P_{D,k}$ for deterministic channels $\{\mathbf{h}_k\}_{k=1}^K$.

The symbols transmitted by the PT are assumed to be independent of one another, and the SUs assume that $\mathbf{s} = \left[\begin{array}{ccc} s[0] & \ldots & s\left[\tilde{M}-1\right] \end{array}\right]^T$ are i.i.d. samples from a zero-mean complex Gaussian random process with $E\left\{\mathbf{s}\mathbf{s}^H\right\} = \sigma_s^2 \mathbf{I}$. We also define $\mathbf{y}_k = \left[\begin{array}{ccc} y_k[0] & \ldots & y_k\left[\tilde{M}-1\right] \end{array}\right]^T$. Since the primary signal has a diagonal covariance matrix and each SU is assumed to have knowledge of its background noise variance, an energy detector (radiometer) is optimal in terms of maximizing the probability of detection P_D for a given false alarm rate P_{FA}. If some or all of these parameters are unknown, then a generalized likelihood ratio test (GLRT) method can be employed [3]. The GLRT methods and other blind techniques usually result in test statistics involving the eigenvalues of the received sample covariance matrix $\mathbf{R}_{k,yy} = \frac{1}{M}\mathbf{y}_k\mathbf{y}_k^H$ [4].

Under the null hypothesis $\mathcal{H}_0$, we see from (4) that $\mathbf{y}_k \sim \mathbb{CN}\left(\mathbf{0}, \sigma_k^2 \mathbf{I}\right)$, whereas under the alternative hypothesis $\mathcal{H}_1$ we have $\mathbf{y}_k \sim \mathbb{CN}\left(\mathbf{0}, \left(\sigma_s^2 \left|\mathbf{h}_k\mathbf{w}\right|^2 + \sigma_k^2\right)\mathbf{I}\right)$. The test statistic for the Neyman-Pearson-optimal energy detector is given by

$$T_k = \sum_{n=0}^{\tilde{M}-1} \left|y_k[n]\right|^2 = \mathbf{y}_k^H \mathbf{y}_k, \tag{5}$$

and the normalized test statistic has a central chi-square distribution with M degrees of freedom under both hypotheses:

$$\begin{array}{ll} \frac{T_k}{\sigma_k^2/2} & \sim \chi_M^2 \quad \text{under } \mathcal{H}_0 \\ \frac{T_k}{\left(\sigma_s^2\left|\mathbf{h}_k\mathbf{w}\right|^2 + \sigma_k^2\right)/2} & \sim \chi_M^2 \quad \text{under } \mathcal{H}_1. \end{array} \tag{6}$$

Hence, for an arbitrary number of samples M and a pre-specified P_{FA}, the detection probability is given by $P_{D,k} = Q_{\chi_M^2}\left(\frac{\lambda_k}{\left(\sigma_s^2\left|\mathbf{h}_k\mathbf{w}\right|^2 + \sigma_k^2\right)/2}\right)$, where $Q_{\chi_M^2}(\cdot)$ is the complementary cdf of the central chi-square distribution, and $\lambda_k = \sigma_k^2 Q_{\chi_M^2}^{-1}(P_{FA})$ is the detection threshold used to distinguish between the hypotheses: $T_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_k$. Since we have an even number of samples M (real and imaginary components of each sample), the false alarm probability follows immediately from the chi-square cdf as [14]

$$P_{FA,k} = e^{-\frac{\lambda_k}{\sigma_k^2}} \sum_{r=0}^{M/2-1} \frac{1}{r!}\left(\frac{\lambda_k}{\sigma_k^2}\right)^r. \tag{7}$$

Similarly, the detection probability is given by

$$P_{D,k} = e^{-\frac{\lambda_k}{\left(\sigma_s^2 |\mathbf{h}_k \mathbf{w}|^2 + \sigma_k^2\right)}} \sum_{r=0}^{M/2-1} \frac{1}{r!} \left(\frac{\lambda_k}{\left(\sigma_s^2 \left|\mathbf{h}_k \mathbf{w}\right|^2 + \sigma_k^2\right)} \right)^r . \tag{8}$$

For i.i.d. Rayleigh fading channels between the PT and the SUs with second-order statistics $\mathbf{C}_k = E\left\{\mathbf{h}_k \mathbf{h}_k^H\right\}$, $P_{D,k}$ in (8) is a function of the random variable $|\mathbf{h}_k \mathbf{w}|^2$, which has an exponential distribution with rate parameter $\frac{1}{\mathbf{w}^H \mathbf{C}_k \mathbf{w}}$ for a deterministic $\mathbf{w}$. Therefore, the average detection probability with Rayleigh fading can be represented as

$$\bar{P}_{D,k} = \frac{1}{\mathbf{w}^H \mathbf{C}_k \mathbf{w}} \int_0^\infty e^{-\frac{\lambda_k}{\left(\sigma_s^2 y + \sigma_k^2\right)}} \sum_{r=0}^{M/2-1} \frac{1}{r!} \left(\frac{\lambda_k}{\left(\sigma_s^2 y + \sigma_k^2\right)} \right)^r e^{-\frac{y}{\mathbf{w}^H \mathbf{C}_k \mathbf{w}}} dy \tag{9}$$

Exchanging the order of the finite summation and integration, then applying a change of variable and [15, Eq. 3.471.9], we arrive at

$$\bar{P}_{D,k} = \frac{\lambda_k}{\sigma_s^2 \mathbf{w}^H \mathbf{C}_k \mathbf{w}} e^{\frac{\sigma_k^2}{\sigma_s^2 \mathbf{w}^H \mathbf{C}_k \mathbf{w}}} \sum_{r=0}^{M/2-1} \left[\frac{2}{r!} \left(\frac{\lambda_k}{\sigma_s^2 \mathbf{w}^H \mathbf{C}_k \mathbf{w}} \right)^{\frac{r-1}{2}} K_{r-1} \left(2 \sqrt{\frac{\lambda_k}{\sigma_s^2 \mathbf{w}^H \mathbf{C}_k \mathbf{w}}} \right) \right], \tag{10}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind with order ν [15]. The false-alarm probability remains unchanged from (7) since it is independent of the fading channel.

C. Cooperative spectrum sensing

We also consider secondary networks employing cooperative spectrum sensing, where the local observations/decisions from all sensors are jointly processed at a designated fusion center to form a global hypothesis test outcome. We consider here a centralized cooperative scheme wherein each SU reports to an external fusion center (FC), which also possesses global CSI of all the SUs [13],[16]-[18]. There are two major categories of cooperative sensing protocols, based on whether the SUs report *hard* information or *soft* information to the FC.

The simplest form of hard information is for each SU to report its local 1-bit binary decision $D_i \in \{1, 0\}$ over an error-free channel to the FC, which then combines the binary reports from all sensors and feeds back the global decision to all SUs. A general hard combining rule is the so-called m-out-of-K counting rule, in which the FC declares the PT to be present if so indicated by at least m of the K total sensors. Note that this rule encompasses the popular “OR”, “AND”, and “Majority” fusion rules by setting $m = 1$, $m = K$, and $m = \left\lceil \frac{K}{2} \right\rceil$, respectively. Recognizing that the m-out-of-K decision rule is equivalent to a Poisson binomial distribution [19] with K independent trials each with a non-uniform probability of success $P_{D,k}$, the detection probability at the FC is

$$P_{D,h}(m) = \frac{\sum_{n=0}^{m} \left\{ \mathrm{e}^{(-j2\pi nm/N+1)} \prod_{k=1}^{K} \left(P_{D,k}\, \mathrm{e}^{(-j2\pi nm/N+1)} + (1 - P_{D,k}) \right) \right\}}{K+1} . \tag{11}$$

The most common approaches for soft information reporting require each SU to send either its unquantized local test statistic T_k or local SNR $\frac{|\mathbf{h}_k \mathbf{w}|^2}{\sigma_k^2}$ to the FC, while alternatives include reporting a scaled copy of the received signal $\mathbf{y}_k$ itself [16]. For both hard and soft information reporting, a variety of models for the SU-FC channels have been investigated, ranging from error-free to fading links with AWGN [18]. Our study of soft information reporting will assume that each SU reports its scaled test statistic $d_k = \omega_k T_k / M$ over a noisy channel to the FC in parallel with the other SUs. The signal from SU k at the FC is $y_{FC,k} = h_{kF} d_k + \nu_k$, $k = 1, \ldots, K$ [18]. Here, the factor ω_k preserves the SU transmit power constraint p_k, h_{kF} is the zero-mean unit-variance channel coefficient to the FC, and ν_k is zero-mean unit-variance complex Gaussian noise. In the limit of a large sample size M, the $K \times 1$ aggregate FC signal $\mathbf{y}_{FC}$ has a Gaussian distribution under both null and alternative hypotheses by the central limit theorem [18]. Subsequently, the likelihood ratio test at the FC assuming knowledge of the effective second hop channel and local SU SNRs reduces to a well-known vector Gaussian detection problem with the corresponding false alarm and detection probabilities [14, Ch. 5],[20, pg. 513].

To conclude this section, we remark that since the PT broadcasts a single information symbol at time t that is simultaneously received by the PR and SUs, the setting is similar to a multicast network, where a common information symbol is sent to a number of different receivers. As such, the ability of the SUs or the FC to detect the presence of the PT signal depends on the local received signal-to-noise ratio (SNR) of the symbol, which for the k^{th} SU is given by

$$\gamma_k = \frac{|\mathbf{h}_k \mathbf{w}|^2}{\sigma_k^2}, \quad k = 1, \ldots, K. \tag{12}$$

III. Primary Receiver Performance Under Secondary Interference

From the perspective of the primary link, a missed detection at any of the SUs when the PT is active leads to interference at the primary receiver. The instantaneous signal-to-interference-plus-noise ratio (SINR) at the primary receiver is

$$\gamma_0 = \frac{|\mathbf{h}_0 \mathbf{w}|^2}{\sigma_0^2 + \sum\limits_{j \in \mathcal{S}} p_j \, |g_j|^2} \tag{13}$$

where $\mathcal{S}$ is the set of SUs that suffer from a missed detection and transmit inadvertently with power p_j and complex channel coefficient $g_j \sim \mathbb{CN}\left(0, \sigma_{g,j}^2\right)$ to the PR. We assume that the sensing duration of the SUs is small compared to the PT transmission interval, such that a missed detection results in virtually instantaneous interference at the PR.

As before, let $P_{D,k}$ represent the detection probability of secondary user k. Dropping the time index for brevity, rewrite the signal at the primary receiver as

$$y_0 = \mathbf{h}_0 \mathbf{x} + \sum_{k=1}^{K} F_k g_k s_k + z_0, \tag{14}$$

where we define the random-valued indicator function F_k as

$$F_k = \begin{cases} 1 & \text{with probability } (1 - P_{D,k}\beta) \\ 0 & \text{with probability } (P_{D,k}\beta) \end{cases}, \tag{15}$$

where $P_{D_k}\beta$ is the probability that SU k is in sensing mode and has detected the presence of the PT, and thus is not producing interference. We are interested in the characteristics of the aggregate SU interference power

$$I_0(\mathbf{w}) = \sum_{k=1}^{K} F_k \, |g_k|^2 \, p_k. \tag{16}$$

Note that (16) depends on $\mathbf{w}$ through the variable F_k, since the ability of the SU to detect the PT depends on $\mathbf{w}$. The distribution of the aggregate interference in interweave networks has been approximated using tools from stochastic geometry in [11], [12] for Poisson point process-distributed SU locations, but this is different from our system model.

Taking the expectation of the SU interference power in (16) with respect to $\{F_k\}_{k=1}^K$ and the SU-PR channels $\{g_k\}_{k=1}^K$ yields

$$I_0'(\mathbf{w}) = \sum_{k=1}^{K} (1 - P_{D,k}\beta) \, \sigma_{g,k}^2 p_k \ , \tag{17}$$

where $\sigma_{g,k}^2$ denotes the variance of the kth SU-PT channel. The PR SINR that can be computed at the PT with partial CSI is thus

$$\gamma_0' = \frac{|\mathbf{h}_0 \mathbf{w}|^2}{\sigma_0^2 + I_0'(\mathbf{w})} \tag{18}$$

where the aggregate interference $I_0'(\mathbf{w})$ is a function of SU parameters β, $\left\{\sigma_{g,k}^2\right\}_{k=1}^K$, $\{p_k\}_{k=1}^K$, and the PT beamformer $\mathbf{w}$ via $P_{D,k}$.

Next, we examine the PR SINR under the assumption of statistical CSI, averaging over the fading channels between the PT and the PR and SUs. Note that $E\left\{F_k | \{\mathbf{h}_k\}_{k=1}^K\right\} = (1 - P_{D,k}\beta)$, and assume that the incoming channels $\{\mathbf{h}_k\}_{k=1}^K$ and outgoing channels $\{g_k\}_{k=1}^K$ of the SUs are independent. Then the expected value of the aggregate interference in (16) averaged over the PT-SU and SU-PR channels is

$$I_0''(\mathbf{w}) = E\{I_0'(\mathbf{w})\} = \sum_{k=1}^K \left(1 - \bar{P}_{D,k}\beta\right)\sigma_{g,k}^2 p_k \ , \tag{19}$$

where $\bar{P}_{D,k}$ is defined in (10).

The average PR SINR that can be computed at the PT when only statistical CSI of all users is available is written as

$$\bar{\gamma}_0 = \frac{\mathbf{w}^H \mathbf{C}_0 \mathbf{w}}{\sigma_0^2 + I_0''(\mathbf{w})}. \tag{20}$$

Having defined the impact of missed detections by the SUs on the performance of the PR, we see that it is in the PT's interest to ensure that the probability of missed detection at the SUs is made as small as possible, or equivalently, that the probability of detection is made as large as possible. To this end, we introduce the paradigm of prescient beamforming in the next section in order to improve the reliability of the primary link.

IV. Prescient Beamforming

A. Motivation

The considerable literature on interweave CR networks generally neglects the role of the primary transmitter and places the entire burden of interference avoidance on the secondary users. Instead, the central problem we analyze in this paper is the following: given some side information in the form of partial or statistical CSI of the SUs, *can the primary transmitter proactively design a reliable data transmission scheme that also minimizes or decreases the probability of interference from the SUs*? In the context of spectrum sensing, this indicates that the PT would like to minimize the probability of missed detection at the CRs. Hence, we seek to design prescient beamforming (PBF) schemes for the primary transmitter, where the term "prescience" derives from the fact that the PT anticipates interference at the PR from SUs due to imperfect spectrum sensing and takes preemptive measures to avoid the same.

The aforementioned delegation of interference avoidance solely to the SUs is due to the original interweave cognitive radio paradigm that was conceived a decade ago. Specifically, primary users were modeled as legacy equipment deployed in conventional infrastructure-based networks that would operate oblivious to the presence of SUs, while SUs were granted cognitive abilities such as spectrum sensing. In such a setting, the expense of upgrading legacy equipment is avoided, and no communication is necessary between the primary and secondary networks. However, in recent models of dynamic spectrum access such as spectrum underlay, there have been proposals for explicit communication between PRs and SUs regarding tolerable and instantaneous interference levels for improved SU power control, for example [23]. Therefore, we feel that it is natural to consider more advanced primary capabilities in an interweave scenario as well. Specifically, we assume that the PT is aware of the noise variances and sensing algorithm in place at the SUs, and has partial channel state information of the PT-SU and SU-PR links.

The proposed PBF schemes are motivated by the simple observation that the detection probabilities of energy or GLRT-based detectors increase monotonically with the received SNR at the SUs for a given false alarm rate P_{FA}. This is clearly seen from (8) for energy detection. For GLRT detection when some or all of the channel and variance parameters are unknown, it is also possible to show that the detection probability is represented by a Gaussian Q-function with the inverse of SNR in the argument, thereby once again $P_{D,k}$ increases with SNR (e.g., see [4, Eqs. (56), (61)]). Furthermore, when the noise variance is not known perfectly at the SUs under energy detection, a 'SNR wall' phenomenon has been observed [21] wherein low-SNR primary signals cannot be reliably detected no matter how many samples are taken ($M \to \infty$). Therefore, the PBF approach attempts to decrease the probability of interference from CRs by increasing their received SNR subject to a desired Quality-of-Service (QoS) for the primary channel if feasible. This approach will also serve to alleviate the SNR wall phenomenon in uncertain noise environments. Naturally, increasing the SNR at the SUs will improve both local and overall detection probabilities for non-cooperative and cooperating sensing, respectively. Therefore, we note that the prescient beamforming techniques proposed in the sequel assuming non-cooperative spectrum sensing can be applied directly without change to the cooperative sensing scenario.

A caveat: it goes without saying that the acquisition of SU CSI needed to enable prescient transmission by the PT incurs additional complexity and overhead costs. However, similar or greater levels of knowledge regarding PR CSI at SUs are routinely assumed in the literature on spectrum underlay and overlay [1],[6]-[10]. Since data is continually transmitted to the PR in PBF, it is also more efficient in terms of throughput compared with the use of a dedicated beaconing phase for the SUs. Furthermore, the cost of any modifications at the primary transmitter is partially offset by the fact that the primary receivers continue to be modeled as oblivious legacy nodes. Finally, we recognize that a dual approach to prescient precoding would be to maintain the primary transmitter as a legacy node and employ multi-user detection/interference cancelation at the primary receivers, although this would be a reactive strategy compared to the proactive schemes in this paper.

In the remainder of this section, we present several prescient transmission schemes that provide a tradeoff between complexity and performance. Each of these schemes can be implemented either with partial or statistical CSI, therefore to avoid repetition we shall illustrate each scheme for one of these CSI assumptions alone.

B. Direct PR SINR Maximization

A first approach to constructing a prescient transmission scheme would be for the PT to directly optimize the PR SINR γ_0' under the transmit power constraint:

$$\begin{aligned} &\max_{\mathbf{w}} \frac{|\mathbf{h}_0 \mathbf{w}|^2}{\sigma_0^2 + I_0'(\mathbf{w})} \\ &s.t.\ \mathbf{w}^H \mathbf{w} \le P/\sigma_s^2\ . \end{aligned} \tag{21}$$

A similar problem can be posed in the case of maximizing the average PR SINR $\bar{\gamma}_0$. Note that the interference term in the PR SINR is a function of the transmit signal itself. While signal-dependent interference is a well-studied problem in radar signal processing, see for example [22], in our case this dependence manifests itself in a much more complicated and non-linear fashion involving exponential terms (and special functions in the case of $\bar{\gamma}_0$). An analytical optimization of (21) over $\mathbf{w}$ appears to be intractable, but one could attempt to find the optimal beamformer using a gradient descent algorithm, recognizing of course that global optimality is not guaranteed due to the non-convexity of the objective function.

The Karush-Kuhn-Tucker (KKT) conditions for a stationary point of (21) can be computed as

$$\nabla_{\mathbf{w}} (\gamma_0') + \mu \mathbf{w} = \mathbf{0} \tag{22}$$

$$\mathbf{w}^H \mathbf{w} - P/\sigma_s^2 \le 0 \tag{23}$$

$$\mu \left(\mathbf{w}^H \mathbf{w} - P/\sigma_s^2 \right) = 0, \quad \mu \ge 0, \tag{24}$$

where the gradient of the PR SINR is given by

$$\nabla_{\mathbf{w}}\left(\gamma_0'\right) = \frac{\left(\mathbf{h}_0^H\mathbf{h}_0\right)\mathbf{w}}{\sigma_0^2+I_0'\left(\mathbf{w}\right)} + \frac{\beta\left|\mathbf{h}_0\mathbf{w}\right|^2\sigma_s^2\sum_{k=1}^{K}\left[\sigma_{g,k}^2 p_k\left(A_k-B_k\right)\right]}{\left(\sigma_0^2+I_0'\left(\mathbf{w}\right)\right)^2} \quad (25)$$

$$\Omega_k\left(\mathbf{w}\right) = \frac{\lambda_k/2}{\sigma_s^2\left|\mathbf{h}_k\mathbf{w}\right|^2+\sigma_k^2} \quad (26)$$

$$A_k = \Omega_k\left(\mathbf{w}\right)e^{-\Omega_k(\mathbf{w})}\mathbf{h}_k^H\mathbf{h}_k\mathbf{w}\sum\nolimits_{r=0}^{M/2-1}\left\{(r!)^{-1}\left(\Omega_k\left(\mathbf{w}\right)\right)^r\right\} \quad (27)$$

$$B_k = e^{-\Omega_k(\mathbf{w})}\frac{\mathbf{h}_k^H\mathbf{h}_k\mathbf{w}}{\sigma_s^2\left|\mathbf{h}_k\mathbf{w}\right|^2+\sigma_k^2}\sum\nolimits_{r=0}^{M/2-1}\left\{(r-1!)^{-1}\left(\Omega_k\left(\mathbf{w}\right)\right)^r\right\}. \quad (28)$$

However, since the problem is non-convex the KKT conditions are merely necessary and not sufficient for optimality.

C. Combined MRT and Multicast Beamforming

While the gradient search algorithm described above returns at least a locally optimal prescient beamformer, it is desirable to investigate designs based on simpler optimization procedures. To this end, consider the following two extreme cases for the choice of $\mathbf{w}$:

- Disregard SUs, focus only on PR: If the PT disregards the presence of the SUs and focuses only on maximizing the signal strength at the PR, the optimal choice for $\mathbf{w}$ is the maximum-ratio transmit beamformer:

$$\mathbf{w}_{MRT} = \frac{\sqrt{P}\mathbf{h}_0^H}{\sigma_s\|\mathbf{h}_0\|}\ . \quad (29)$$

- Disregard PR, focus only on SUs: At this extreme, the PT ignores the PR and focuses only on improving the signal strength at the SUs (particularly those that could produce the most interference at the PR). This is similar to a multicast (MC) downlink scenario, where priority is given to certain key users. A reasonable choice for the transmit beamformer in this case would maximize the weighted average of the SNRs at the SUs:

$$\mathbf{w}_{MC} = \arg\max_{\mathbf{w}}\sum_{k=1}^{K}p_k\sigma_{g,k}^2|\mathbf{h}_k\mathbf{w}|^2\ , \quad (30)$$

  where the weight $p_k\sigma_{g,k}^2$ measures the interference impact of the kth SU at the PR. It is easy to see that the solution to (30) is given by the dominant singular vector of $\mathbf{H}_S^H\mathbf{\Sigma}_g\mathbf{H}_S^H$ scaled by $\sqrt{P/\sigma_s^2}$, where $\mathbf{H}_S = \left[\begin{array}{ccc}\mathbf{h}_1^T & \ldots & \mathbf{h}_K^T\end{array}\right]^T$ and $\mathbf{\Sigma}_g$ is a diagonal matrix with entries $p_k\sigma_{g,k}^2, k=1,\cdots,K$.

Given that the prescient beamforming objective is to balance these two competing goals, a sensible approach would be to choose $\mathbf{w}$ as some linear combination of the solutions:

$$\mathbf{w} = \alpha\mathbf{w}_{MRT} + (1-\alpha)\,\mathbf{w}_{MC} \qquad 0 \le \alpha \le 1\ , \quad (31)$$

where the optimal value of $\alpha \in [0,1]$ can be found by a simple line search. We will see later in the simulations that this approach performs similarly to the gradient search for maximizing the PR SINR.

D. Algorithms based on Convex Optimization

Here we investigate alternative cost functions for the prescient beamforming problem that lead to solutions based on convex optimization. We treat two cases: (1) maximize the signal power delievered to the PR subject to a received signal power constraint at the SUs, and (2) maximize the worst-case probability of detection at the SUs while delivering a desired signal power at the PR. We will present these problems under the statistical CSI assumption, but they can easily be formulated for the partial CSI case.

The first problem can be formulated as follows:

$$\begin{aligned} &\max_{\mathbf{w}} E_{\mathbf{h}_0}\{|\mathbf{h}_0\mathbf{w}|^2\} \\ &s.t.\ E_{\mathbf{h}_k}\{|\mathbf{h}_k\mathbf{w}|^2\} \geq \eta_k,\ k=1,\ldots,K \\ &\mathrm{Tr}\left(\mathbf{w}\mathbf{w}^H\right) \leq P/\sigma_s^2\ , \end{aligned} \tag{32}$$

where η_k is the desired threshold on the SNR at SU k, which in turn corresponds to a desired detection probability $P_{D,k}$. If we define $\mathbf{J} \triangleq \mathbf{w}\mathbf{w}^H$ and $\mathbf{C}_k = E\{\mathbf{h}_k^H\mathbf{h}_k\}$, we can rewrite the problem in (32) as a relaxed *max-PR power* semidefinite program (SDP):

$$\begin{aligned} &\max_{\mathbf{J}} \mathrm{Tr}\left(\mathbf{J}\mathbf{C_0}\right) \\ &s.t.\ \mathrm{Tr}\left(\mathbf{J}\right) \leq P/\sigma_s^2 \\ &\mathrm{Tr}\left(\mathbf{J}\mathbf{C}_k\right) \geq \eta_k, \quad k=1,\ldots,K \\ &\mathbf{J} \succeq 0 \end{aligned} \tag{33}$$

which leads to an approximate solution due to the relaxation of the rank-1 constraint on $\mathbf{J}$. Since the objective function and all constraints are convex, this SDP can be solved efficiently using interior-point methods. If the computed result is not rank-1, then well-known randomization techniques can be applied to obtain an appropriate solution [10].

The second convex problem can be stated as

$$\begin{aligned} &\max_{\mathbf{J}} \min_{k} \mathrm{Tr}\left(\mathbf{J}\mathbf{C}_k\right) \\ &s.t.\ \ \mathrm{Tr}\left(\mathbf{J}\right) \leq P/\sigma_s^2 \\ &\mathbf{J} \succeq 0 \\ &\mathrm{Tr}\left(\mathbf{J}\mathbf{C_0}\right) \geq \sigma_0^2\eta_{0,min} \end{aligned} \tag{34}$$

After a change of variable, this can also be formulated as a semidefinite program, which we denote as the *max-min SU power* SDP.

E. Performance Upper Bound

In this section we develop information-theoretic bounds on the capacity and consequently the maximum achievable SINR of the primary channel, obtained possibly by non-linear prescient precoding techniques. Conventional mutual information results for a fading broadcast channel are not applicable due to the presence of signal-dependent interference and since the K SUs do not decode the primary signal.

If we assume the primary transmitter has instantaneous CSI of *all* links, as well as additional side information (SI) in the form of prior knowledge of the secondary signals $\{s_k\}_{k=1}^K$, then we can develop an omniscient (genie) upper bound for the primary channel capacity. In practice, the SU signals can be known at the PT in certain specialized scenarios, e.g., where the SUs opportunistically relay the primary signal itself or retransmit their own data via ARQ, and such assumptions have also been made in the literature on spectrum overlay CRs [1], [24], [28]. Revisiting the received primary signal in (14), we can draw connections to the scalar dirty paper channel since the SU interference is *partially* known to the PT:

$$y_0 = \mathbf{h}_0\mathbf{x} + \mathbf{F}\mathbf{G}\tilde{\mathbf{s}} + z_0 \tag{35}$$

where $\mathbf{F} = \left[\begin{array}{ccc} F_1 & \dots & F_K \end{array} \right]$ is the collection of the SU indicator variables, $\mathbf{G} = \mathrm{diag}\left[g_1, \dots, g_K\right]$, $\tilde{\mathbf{s}} = \left[\begin{array}{ccc} s_1 & \dots & s_K \end{array} \right]^T$, and we define $\mathbf{P}_K \triangleq E\left\{\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H\right\} = \mathrm{diag}\left[p_1, \dots, p_K\right]$. Here, the interference $\mathbf{G}\tilde{\mathbf{s}}$ is the side information known to the PT but not the receiver, and it is multiplied by a random vector $\mathbf{F}$ which has a known joint probability distribution since it comprises K independent non-identically distributed Bernoulli random variables (cf. (15)). This signal model differs from existing work on robust dirty paper coding, e.g. [25]-[29], that have considered the case where the known interference is multiplied by a random variable that represents channel fading.

The capacity achieved with dirty paper coding depends on whether the side information or state $\mathbf{G}\tilde{\mathbf{s}}$ is known causally or non-causally to the PT. If the codeword duration is less than the minimal SU-to-PR channel coherence time, we can assume the interference term $\mathbf{G}\tilde{\mathbf{s}}$ is constant for each code symbol and is known non-causally. For brevity, let the state $\mathbf{G}\tilde{\mathbf{s}}$ be represented by $\mathbf{T}$ with realization $\mathbf{t}$, and let $\mathbf{X}$ denote the random variable counterpart of the channel input $\mathbf{x}$, with covariance $E\left\{\mathbf{X}\mathbf{X}^H\right\} = \mathbf{\Sigma}_X$. When the PT-to-SU channels $\{\mathbf{h}_k\}_{k=1}^K$ are time-varying and the realizations of $\mathbf{F}$ are known to the primary receiver but not the transmitter, the ergodic primary channel capacity is given by the modified Gel'fand-Pinsker expression [26], [27]

$$\max_{\substack{p(\mathbf{u}|\mathbf{t}), p(\mathbf{x}|\mathbf{u},\mathbf{t}) \\ \mathrm{Tr}(\mathbf{\Sigma}_X) \le P}} E_{\mathbf{F}}\left\{I\left(\mathbf{U}; Y, \mathbf{F}\right) - I\left(\mathbf{U}; \mathbf{T}\right)\right\} \tag{36}$$

where $I\left(\cdot;\cdot\right)$ represents mutual information, Y is the random variable counterpart of y_0, $\mathbf{U}$ is an auxiliary random variable such that $\mathbf{U} \to \left(\mathbf{x}, \mathbf{T}\right) \to \left(y_0, \mathbf{F}\right)$ forms a Markov chain, and $p\left(\mathbf{x}|\mathbf{u},\mathbf{t}\right) = 1$ if $\mathbf{x} = f\left(\mathbf{u},\mathbf{t}\right)$ for some deterministic function $f\left(\mathbf{u},\mathbf{t}\right)$.

For the above scenario, an achievable rate is given by selecting the channel input $\mathbf{X}$ from a zero-mean complex Gaussian distribution. Furthermore, following a *linear assignment* strategy, the auxiliary variable $\mathbf{U}$ is chosen as $\mathbf{U} = \mathbf{X} + \mathbf{\Phi}\mathbf{T}$, or equivalently $f\left(\mathbf{u},\mathbf{t}\right) = \mathbf{u} - \mathbf{\Phi}\mathbf{t}$, where the $\left(N \times K\right)$ matrix $\mathbf{\Phi}$ is the dirty paper coding *inflation factor* [29]. Given these choices and recognizing that $\mathbf{U}$ and $\mathbf{F}$ are independent, we can evaluate the argument inside the expectation of (36) as

$$\begin{aligned} R_{DP} &= I\left(\mathbf{U}; Y|\mathbf{F}\right) - I\left(\mathbf{U};\mathbf{T}\right) && (37)\\ &= h\left(\mathbf{U}|\mathbf{F}\right) - h\left(\mathbf{U}|Y,\mathbf{F}\right) - h\left(\mathbf{U}\right) + h\left(\mathbf{U}|\mathbf{T}\right) && (38)\\ &= h\left(Y\right) - h\left(\mathbf{U}, Y\right) - h\left(\mathbf{T}\right) + h\left(\mathbf{U},\mathbf{T}\right) && (39)\\ &= \log\left(\pi e \sigma_y^2\right) - \log\left(\left|\pi e \mathbf{\Sigma}_{U,Y}\right|\right) - \log\left(\left|\pi e \mathbf{\Sigma}_T\right|\right) + \log\left(\left|\pi e \mathbf{\Sigma}_{U,T}\right|\right) && (40) \end{aligned}$$

where $h\left(\cdot\right)$ indicates differential entropy. The covariances required to compute the achievable rate are

$$\begin{aligned} &\sigma_y^2 = \mathbf{h}_0\mathbf{\Sigma}_X\mathbf{h}_0^H + \mathbf{F}\mathbf{G}\mathbf{P}_K\mathbf{G}^H\mathbf{F}^H + \sigma_0^2 \\ &\mathbf{\Sigma}_T = \mathbf{G}\mathbf{P}_K\mathbf{G}^H \\ &\mathbf{\Sigma}_{U,T} = \left[\begin{array}{cc} \mathbf{\Sigma}_X + \mathbf{\Phi}\mathbf{\Sigma}_T\mathbf{\Phi}^H & \mathbf{\Phi}\mathbf{\Sigma}_T \\ \mathbf{\Sigma}_T\mathbf{\Phi}^H & \mathbf{\Sigma}_T \end{array} \right] \\ &\mathbf{\Sigma}_{U,Y} = \left[\begin{array}{cc} \mathbf{\Sigma}_X + \mathbf{\Phi}\mathbf{\Sigma}_T\mathbf{\Phi}^H & \mathbf{\Sigma}_X\mathbf{h}_0^H + \mathbf{\Phi}\mathbf{\Sigma}_T\mathbf{F}^H \\ \mathbf{h}_0\mathbf{\Sigma}_X + \mathbf{F}\mathbf{\Sigma}_T\mathbf{\Phi}^H & \sigma_y^2 \end{array} \right]. \end{aligned} \tag{41}$$

Finally, the ergodic dirty-paper rate is given by $E_{\mathbf{F}}\left\{R_{DP}\right\}$, and therefore an upper bound on the primary SINR is $\gamma_{0,UB} = 2^{E_{\mathbf{F}}\{R_{DP}\}} - 1$. Note that this expectation exists and is finite since $\mathbf{F}$ is drawn from a discrete-valued distribution. Numerical algorithms to compute the inflation factor $\mathbf{\Phi}$ and $\mathbf{\Sigma}_X$ can be found in [29]. On the other hand, if $\mathbf{F}$ is 'quasi-static' over the codeword duration then an outage probability metric is more appropriate since the ergodic rate is zero; see [27], [28] for discussions on robust dirty paper coding in this context.

Recall that the dirty paper rate relies on random coding arguments with Gaussian inputs. We can attempt to construct a practical non-linear transmission technique based on Tomlinson-Harashima precoding (THP), which is well known as a low-complexity but suboptimal scalar implementation of dirty paper coding

[31], [32]. To illustrate the concept, consider the case where the PT has complete side information of all channels, secondary signals $\tilde{s}$, and the realization of $\mathbf{F}$. Let the primary data symbol s_0 be drawn from an M-ary modulation alphabet. The effective primary data symbol is generated by pre-subtracting the known interference

$$u_0 = m_\tau \left(s_0 - \mathbf{F}\mathbf{G}\tilde{\mathbf{s}} \right), \tag{42}$$

where $m_\tau(z) = z - \lfloor 0.5 + \mathrm{Re}(z)/\tau \rfloor \tau - j \lfloor 0.5 + \mathrm{Im}(z)/\tau \rfloor$ is the modulo operator with respect to τ which is dependent on the constellation size M, and the transmit signal is given by $\mathbf{x} = \mathbf{w}u_0$. At the receiver, a scaling factor and a second modulo operation are applied prior to data detection. In this complete-side-information scenario, the dirty paper capacity is the same as the primary channel capacity without any SU interference; however, the classic THP scheme suffers from unavoidable shaping and modulo losses even though the interference is completely removed [31].

If only the distribution of $\mathbf{F}$ is known to the primary transmitter, the SU interference can no longer be completely pre-canceled by the PT using the THP scheme. A naïve implementation of THP in this scenario would be to round the mean of $\mathbf{F}$ to integer values and use the resulting quantized $\mathbf{F}$ for interference presubtraction as in (42). A more robust strategy that exploits the statistics of $\mathbf{F}$ to minimize the mean-square error of the data symbol at the primary receiver is presented in [32], but comes at the cost of increased complexity as well as the need for feedback of the residual interference covariance from the receiver.

V. Simulation Results

In this section, we present the results of several numerical experiments to verify the improvement in primary link performance with prescient beamforming. To avoid repetition, unless specified otherwise, all results in this section are based on the partial CSI model with instantaneous CSI of the PT-PR and PT-SU links, and only statistical CSI of the SU-PR links available to the primary transmitter. Each channel realization for all terminals is drawn from a zero-mean circularly symmetric complex Gaussian distribution, and all results are averaged over 1000 channel realizations. The AWGN variance at all receivers is assumed to be unity, i.e., $\sigma_k^2 = 1 \, \forall k$. For every non-cooperative sensing scenario we assume a sensing probability of $\beta = 0.9$, false alarm rate $P_{FA} = 10^{-3}$, and sample size of $M = 4$ used by the SUs for detection. The gradient-PBF algorithms are run 5 times for each set of channel realizations with four random initializations and an initialization based on the naïve MRT precoder to reduce the likelihood of a local maximum; the best-performing precoding solution is chosen as the result. The prescient SDP schemes are implemented using the cvx Matlab toolbox. If, for a given set of channels, the initial SNR targets $\eta_k = 2P/3$ are not feasible, they are decreased by 1% and the solver is rerun.

We first examine the energy detection receiver-operating-characteristic for PBF compared to MRT transmission with a single primary receiver, for both non-cooperative and cooperative (hard reporting using OR fusion; soft reporting over parallel channels) spectrum sensing in Fig. 2. The primary transmitter has $N = 4$ antennas, and its transmit power is fixed at $P = 5dB$. There are $K = 4$ secondary users, and each SU transmits at $p_k = 20dB$. For the cooperative sensing systems, any gains from PBF are overshadowed by the spatial diversity gain in both schemes due to cooperation. In contrast, we observe that PBF provides a significant improvement in non-cooperative energy detection performance for the entire range of P_{FA}.

In Fig. 3, we compare the primary SINR for prescient versus naïve transmission as a function of the primary transmit power. Here, the fixed parameters are $N = 4, K = 5, p_k = 20dB$. The exploitation of SU CSI affords a performance increase of around $3dB$ at low to moderate SNRs. The combined MRT-MC beamformer (referred to as "linear combination" in the plot) of (31) has a negligible SINR loss compared to the direct approach in (21). While the gap between the DPC upper bound and PBF is significant, note that the DPC bound also relies on non-causal knowledge of the SU signals in addition to complete CSI. We see that as P increases, the PBF SINR converges more quickly than MRT to the partial-SI DPC upper bound, which is expected due to the negligible probabilities of missed detection in this regime.

Fig. 4 displays the aggregate SU interference at the primary receiver for the various beamforming schemes, with the same parameters as in Fig. 3. While the gradient-PBF virtually eliminates SU interference due to missed detections at high SNRs, the corresponding increase in PR SINR is not as significant (cf. Fig. 3), since the PBF transmission also degrades the desired signal power when compared to MRT.

We compare the max-min SU power SDP and MRT schemes as a function of the number of primary transmit antennas N in Fig. 5. The PR signal power thresholds are fixed at $5dB$ with transmit power $P = 25dB$. In all instances, the corresponding worst-case SU detection probability is substantially improved under PBF (up to 25%) compared with the naïve primary beamforming strategy. More importantly, merely increasing the number of primary transmit antennas is not sufficient to improve $P_{D,k}$ for the naïve schemes, while the PBF strategies exploit the transmit degrees-of-freedom more efficiently.

Next, we consider the statistical CSI model where the PT possesses information only of the second-order statistics of its outgoing channels. We generate the channel covariances as $\mathbf{C}_k = (1-v)\mathbf{a}_k\mathbf{a}_k^H + v\mathbf{I}$ for each k, where $\mathbf{a}_k$ is the steering vector for a uniform linear array with a given angle-of-arrival, and $0 \leq v \leq 1$. Note that when $v = 1$, the PT assumes the channels are spatially uncorrelated and is essentially choosing random isotropic beamformers. For the special case of a single SU with a near-orthogonal steering vector compared to the PR, the average primary SINR is shown in Fig. 6 for $N = 5, P = 20dB, p_k = 30dB$ as a function of parameter v, and PBF is seen to provide a substantial gain over MRT up to $v = 0.5$. As the precision of the statistical CSI decreases with increasing v, the PBF and MRT schemes coincide as expected.

VI. Conclusion

In this work we considered the novel problem of linear precoding by the primary transmitter to increase the detection probability at spectrum-sensing cognitive radios, and thereby decreasing the inadvertent interference at the primary receiver due to imperfect spectrum sensing. We devised a variety of prescient beamforming schemes with differing complexities that preemptively mitigate secondary interference and are applicable to both non-cooperative and cooperative spectrum sensing. We also computed information-theoretic upper bounds on the maximal achievable SINR based on dirty paper coding schemes. Numerical results demonstrate that the primary link performance is improved under the proposed prescient beamforming methods. In forthcoming work, we are investigating the extension to the case of multiple primary receivers and corresponding prescient precoding strategies.

References

[1] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, pp. 894-914, May 2009.

[2] J. F.-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2137-2146, Jul. 2010.

[3] R. Zhang, T. J. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84-88, Jan. 2010.

[4] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814-823, Feb. 2010.

[5] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Tech.*, vol. 59, no. 4, pp. 1791-1800, May 2010.

[6] R. Zhang and Y. C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 88-102, Feb. 2008.

[7] A. Tajer, N. Prasad, and X. Wang, "Beamforming and rate allocation in MISO cognitive radio networks," *IEEE Trans. Sig. Proc.*, vol. 58, no. 1, pp. 362-377, Jan. 2010.

[8] L. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel state information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4143-4153, Aug. 2009.

[9] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Tech.*, vol. 59, no. 6, pp. 2852-2867, Jul. 2010.

[10] K. T. Phan, S. A. Vorobyov, N. D. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Trans. Sig. Proc.*, vol. 57, no. 6, pp. 2323-2335, Jun. 2009.

[11] A. Ghasemi and E. S. Sousa, "Interference aggregation in spectrum-sensing cognitive wireless networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 41-56, Feb. 2008.

[12] A. Rabbachin, T. Q. S. Quek, H. Shin, and M. Z. Win, "Cognitive network interference," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 480-493, Feb. 2011.

[13] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4502-4507, Nov. 2008.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing Vol. II: Detection Theory*. Prentice Hall, 1998.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 7th ed., 2007.

[16] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," to appear, *IEEE Trans. Wireless Commun.*, 2011.

[17] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 2840, Feb. 2008.

[18] G. Xiong, S. Kishore, and A. Yener, "On performance evaluation of cooperative spectrum sensing in cognitive radio networks," in *Proc. CISS*, pp. 1-6, Princeton, NJ, Mar. 2010.

[19] M. Fernandez and S. Williams, "Closed-form expression for the Poisson-Binomial probability density function," *IEEE Trans. Aero. and Electr. Sys.*, vol. 46, no. 2, pp. 803-817, Apr. 2010.

[20] T. K. Moon, and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[21] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 417, Feb. 2008.

[22] S. U. Pillai, H. S. Oh, D. C. Youla, and J. R. Guerci, "Optimum transmit-receiver design in the presence of signal-dependent interference and channel noise," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 577-584, Mar. 2000.

[23] S. K. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3300-3310, Jun. 2009.

[24] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," IEEE Transactions on Information Theory, vol. 55, no. 9, pp. 3945-3958, Sep. 2009.

[25] A. Khina and U. Erez, "On the robustness of dirty paper coding," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1437-1446, May 2010.

[26] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Prob. Cont. and Inf. Theory*, vol. 9, no. 1, pp. 19-31, 1980.

[27] P.-H. Lin, S.-C. Lin, C.-P. Lee, and H.-J. Su, "Cognitive radio with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3402-3413, Nov. 2010.

[28] P. Mitran, N. Devroye, and V. Tarokh, "On compound channels with side information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1745-1755, Apr. 2006.

[29] C. S. Vaze and M. K. Varanasi, "Dirty paper coding for the MIMO cognitive radio channel with imperfect CSIT," in *Proc. IEEE ISIT*, pp. 2532-2536, Seoul, S. Korea, Jul. 2009.

[30] A. Bennatan and D. Burshtein, "On the fading paper achievable region of the fading MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 100-115, Jan. 2008.

[31] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305-1316, Jul. 2004.

[32] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima precoding for the wireless broadcast channel," *IEEE Trans. Sig. Proc.*, vol. 55, no. 2, pp. 631-644, Feb. 2007.

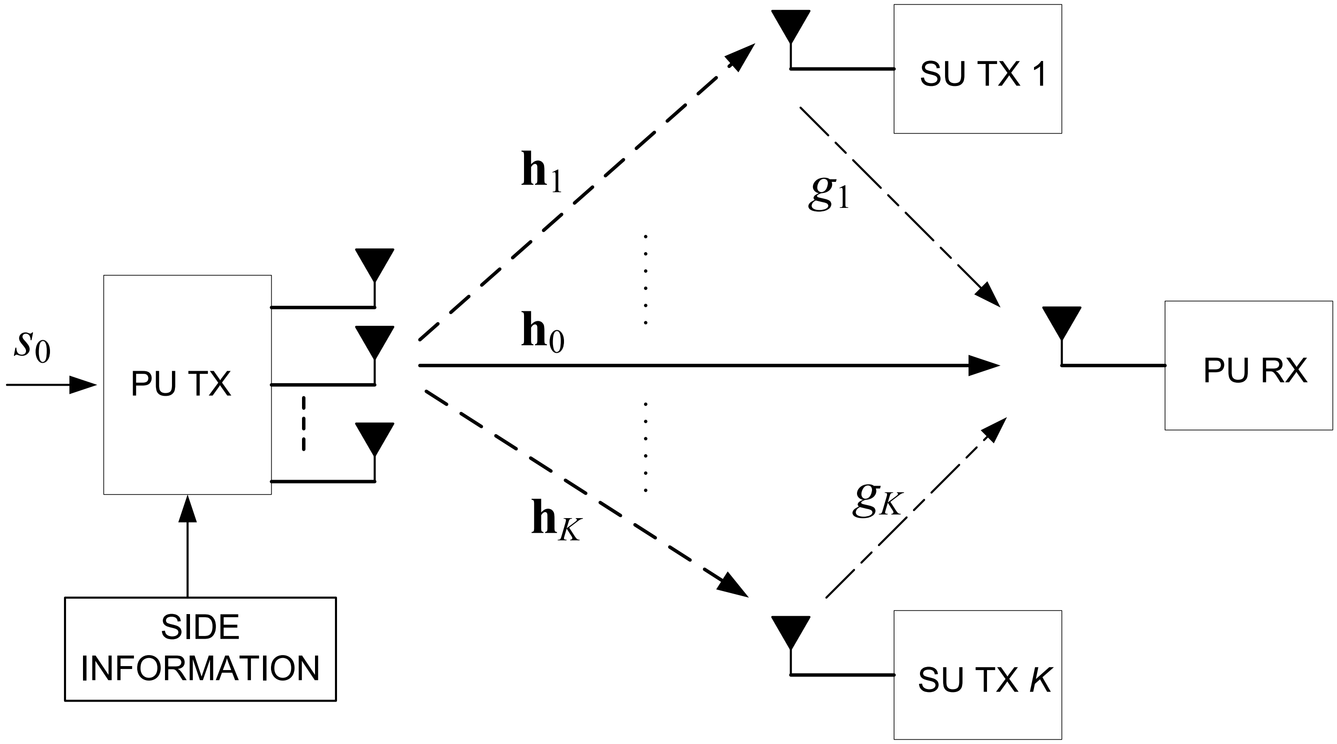

Fig. 1. Cognitive radio network with a multi-antenna primary transmitter, a single primary receiver, and K secondary transmitters. The secondary receivers are not shown for clarity.

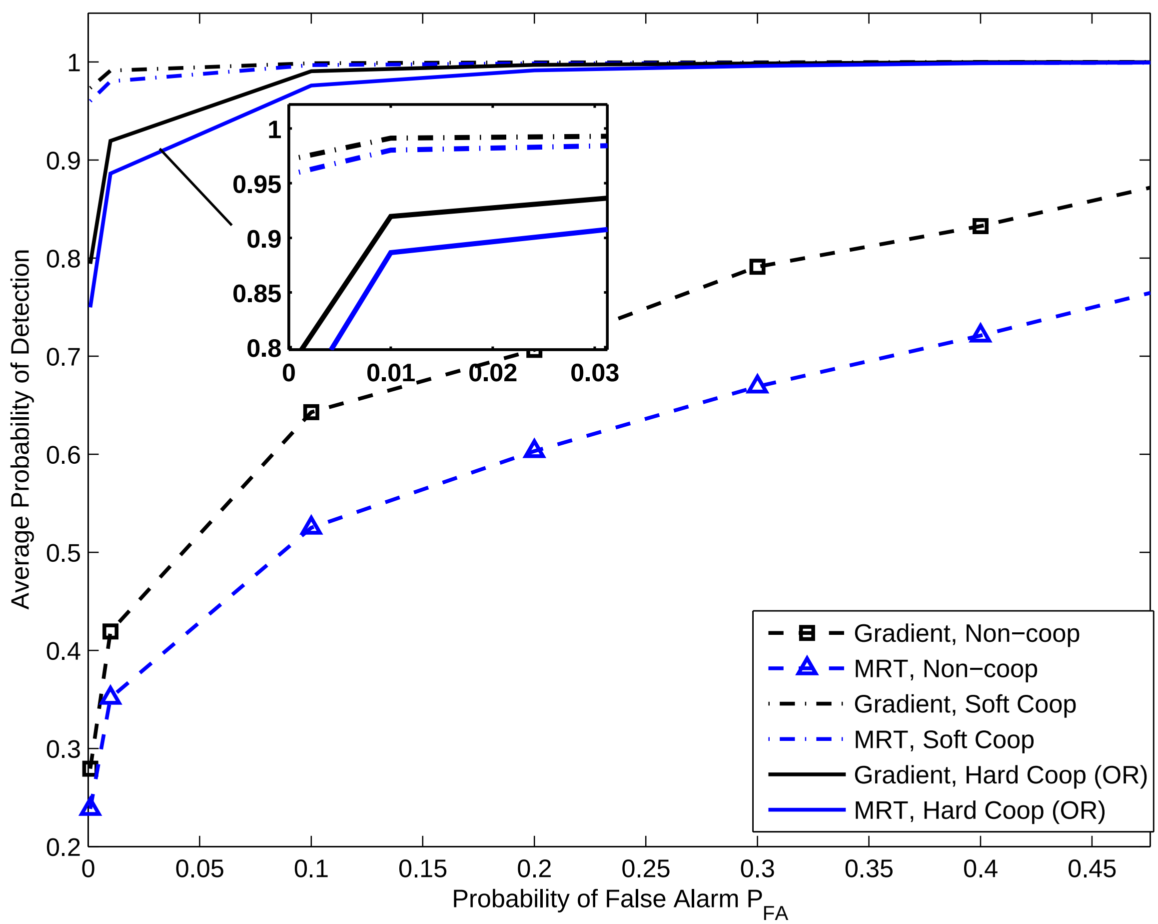

Fig. 2. ROC curve for energy detection comparing prescient beamforming with maximum ratio transmission, $N = 4, P = 5dB$.

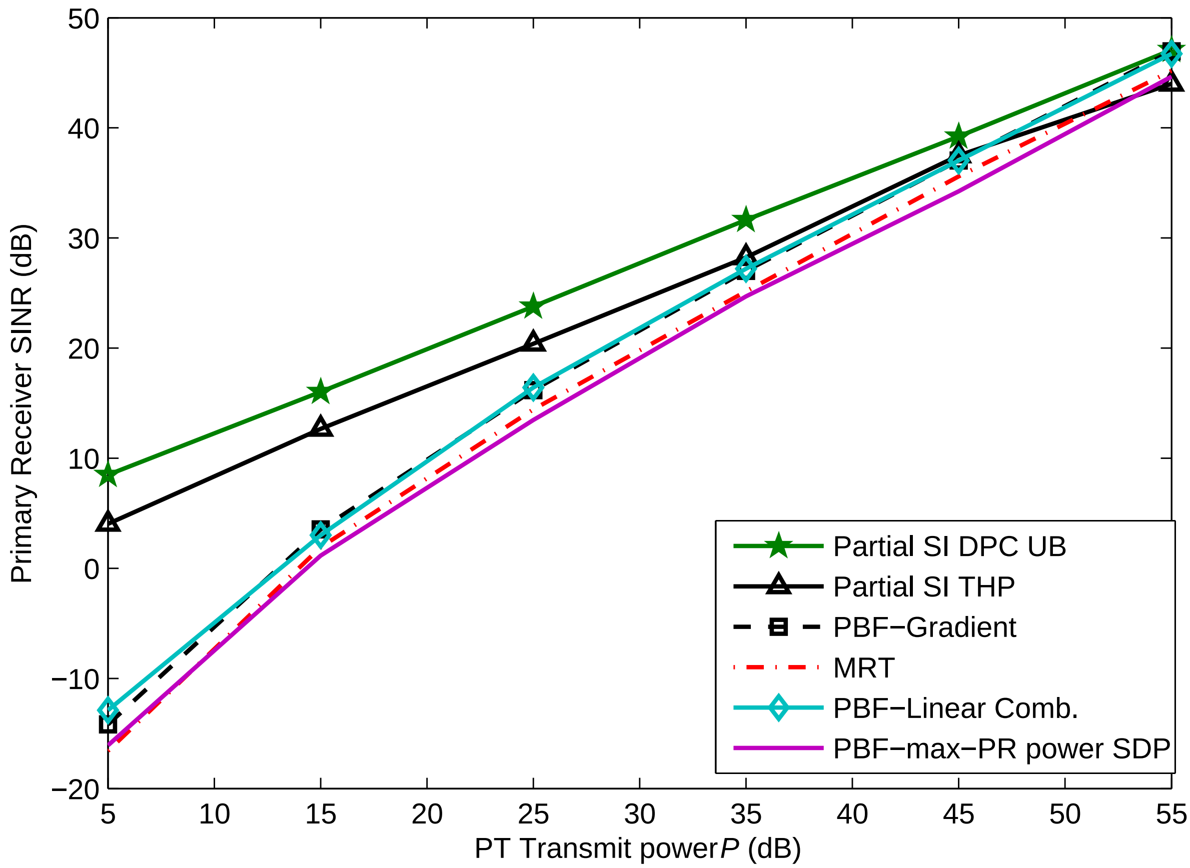

Fig. 3. Primary receiver SINR for non-cooperative spectrum sensing with dirty paper coding, Tomlinson-Harashima precoding, PBF and MRT transmission. SU transmit power $p_k = 20dB$, $K = 5, N = 4, M = 4$.

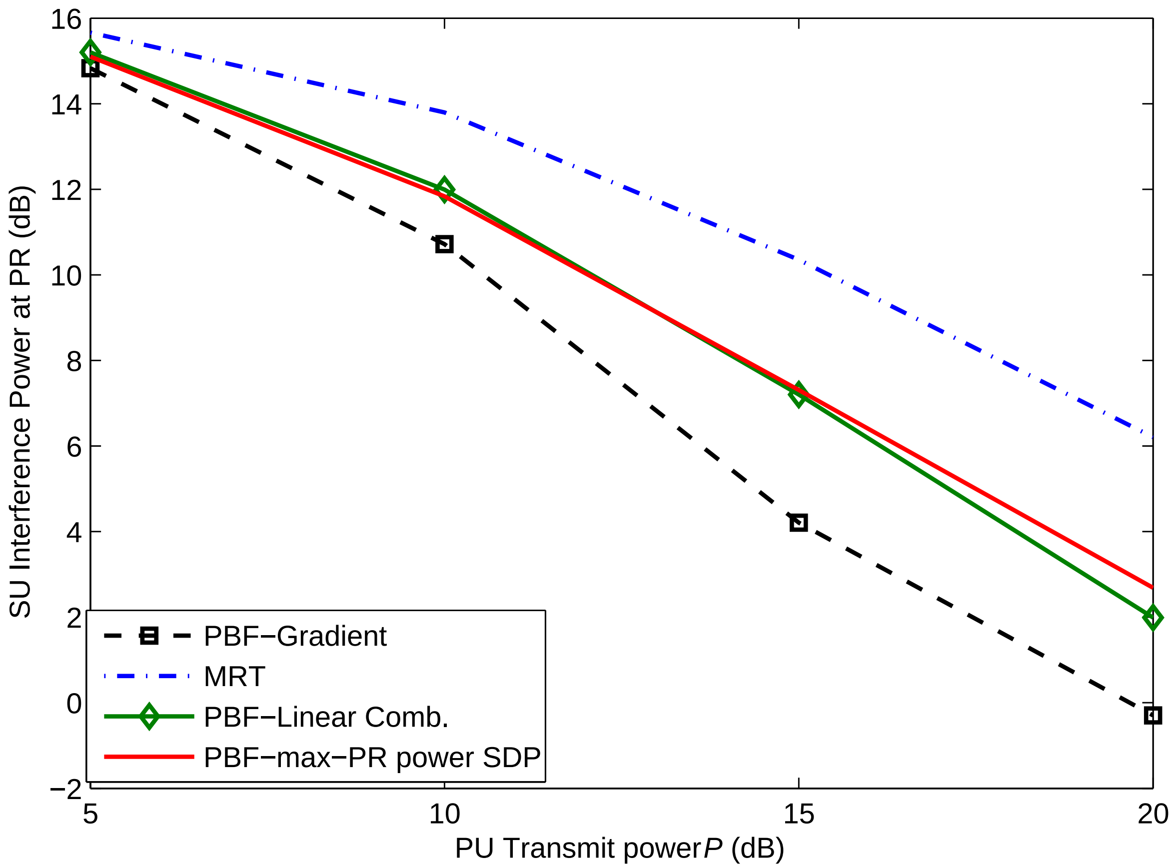

Fig. 4. SU interference power at primary receiver for PBF and MRT schemes with non-cooperative spectrum sensing, $p_k = 20dB, K = 5, N = 4$.

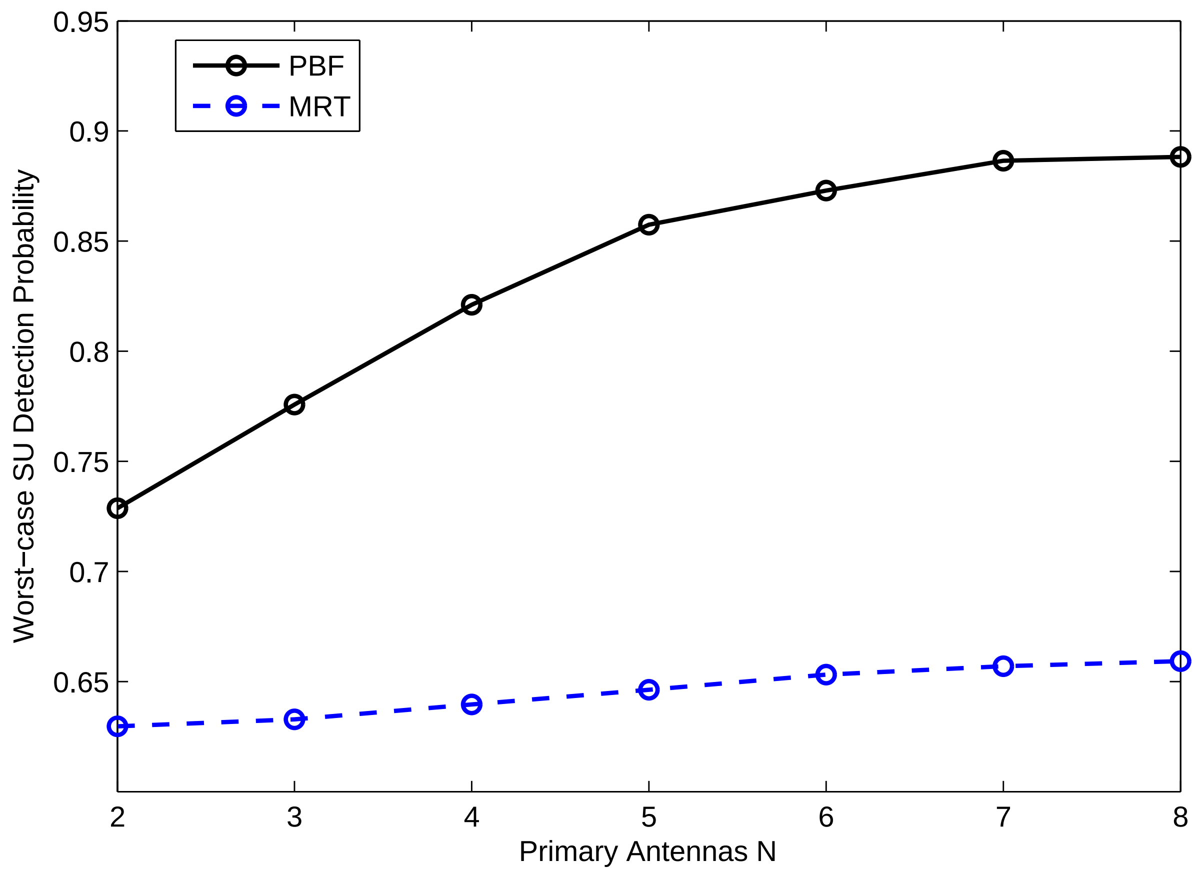

Fig. 5. Worst-case SU detection probabilities based on max-min SU power SDPs.

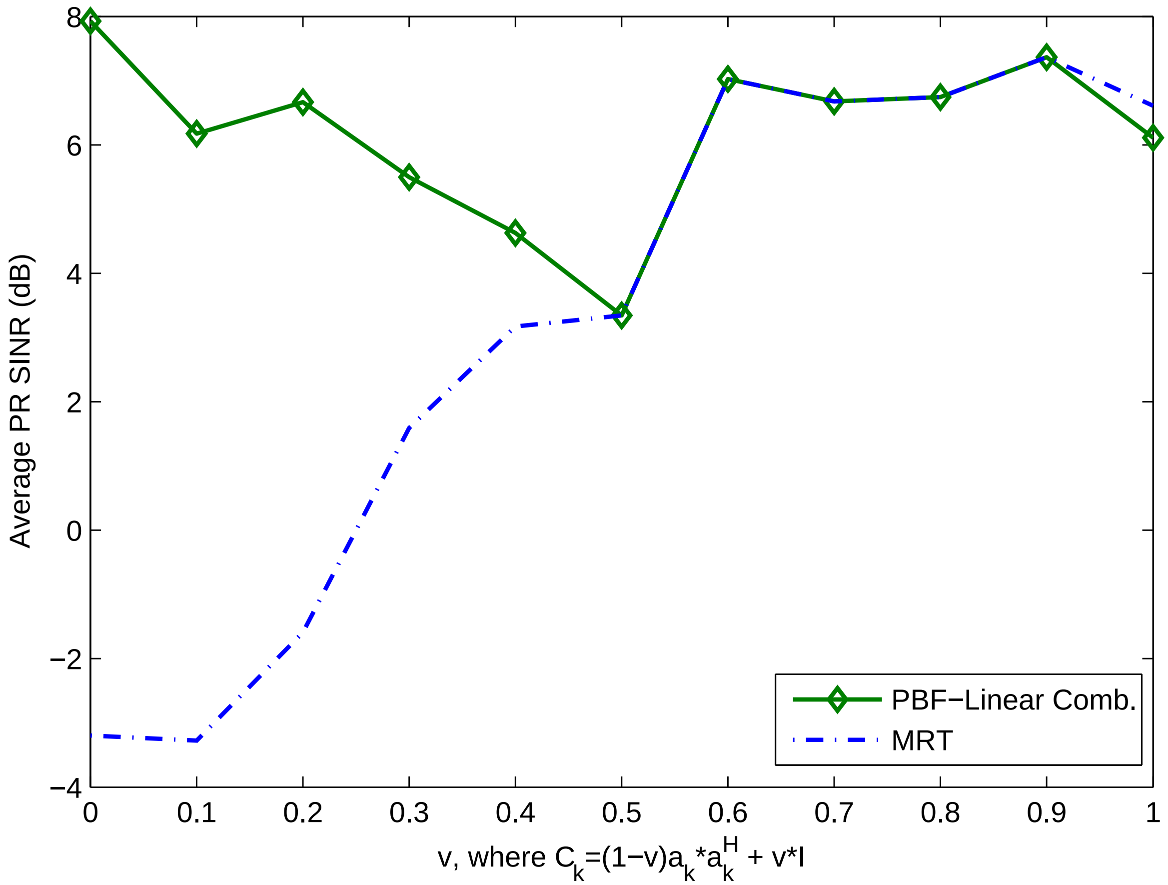

Fig. 6. Average PR SINR versus parameter v for $N = 5, K = 1, p_k = 30dB$.