Cognitive Radio: An Information-Theoretic Perspective

Aleksandar Jovičić and Pramod Viswanath *†

May 8, 2006

Abstract

Cognitive radios have been proposed as a means to implement efficient reuse of the licensed spectrum. The key feature of a cognitive radio is its ability to recognize the primary (licensed) user and adapt its communication strategy to minimize the interference that it generates. We consider a communication scenario in which the primary and the cognitive user wish to communicate to different receivers, subject to mutual interference. Modeling the cognitive radio as a transmitter with side-information about the primary transmission, we characterize the largest rate at which the cognitive radio can reliably communicate under the constraint that (i) *no interference* is created for the primary user, and (ii) the primary encoder-decoder pair is oblivious to the presence of the cognitive radio.

1 Introduction

Observing a severe under-utilization of the licensed spectrum, the FCC has recently recommended [7, 8] that significantly greater spectral efficiency could be realized by deploying wireless devices that can coexist with the incumbent licensed (primary) users, generating minimal interference while somehow taking advantage of the available resources. Such devices could, for instance, form real-time secondary markets [14] for the licensed spectrum holders of a cellular network or even, potentially, allow a complete secondary system to simultaneously operate in the same frequency band as the primary.

*A. Jovičić and P. Viswanath are with the department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. Email: {jovicic,pramodv}@uiuc.edu

†This research was supported in part by the National Science Foundation under grant CCR-0312413 and a grant from Motorola Inc. as part of the Motorola Center for Communication.

The characteristic feature of these *cognitive radios* would be their ability to *recognize* their communication environment and *adapt* the parameters of their communication scheme to maximize the quality of service for the secondary users while minimizing the interference to the primary users.

In this paper, we study the fundamental limits of performance of wireless networks endowed with cognitive radios. In particular, in order to understand the ultimate system-wide benefits of the cognitive nature of such devices, we assume that the cognitive radio has non-causal knowledge of the codeword of the primary user in its vicinity[1]; in this, we are motivated by the model proposed in [6]. We address the following fundamental question:

What is the largest rate that the cognitive radio can achieve under the constraint that

(i) *it generates no interference for the primary user in its vicinity, and*

(ii) *the primary receiver uses a single-user decoder, just as it would in the absence of the cognitive radio?*

We will refer to these two imperative constraints as the *coexistence conditions* that a cognitive secondary system must satisfy.

Of central interest to us is the communication scenario illustrated in Fig. 1: The primary user wishes to communicate to the primary base-station B_p. In its vicinity is a secondary user equipped with a cognitive radio that wishes to transmit to the secondary base-station B_s. We assume that the cognitive radio has obtained the message of the primary user. The received signal-to-noise ratio of the cognitive radio's transmission at the secondary base-station is denoted by SNR. The transmission of the cognitive radio is also received at B_p, and the signal-to-noise ratio of this interfering signal is denoted by INR (interference-to-noise ratio). If the cognitive user is close to B_p, INR could potentially be large.

Our main result is the characterization of the largest rate at which the cognitive radio can reliably communicate with its receiver B_s under the coexistence conditions and in the "low-interference-gain" regime in which $\mathsf{INR} \leq \mathsf{SNR}$. This regime is of practical interest since it models the realistic scenario in which the cognitive radio is closer to B_s than to B_p. Moreover, we show that the capacity achieving strategy is for the cognitive radio to perform *precoding* for the primary users's codeword and transmit over the *same* time-frequency slot as that used by the primary radio.

To prove our main result, we allow the primary and secondary systems to *cooperate*

[1]Note that this does not imply that the cognitive user can decode the *information* that the primary user is communicating since there are secure encryption protocols running at the application layer. The decoded codeword is a meaningless stream of bits for the cognitive user.

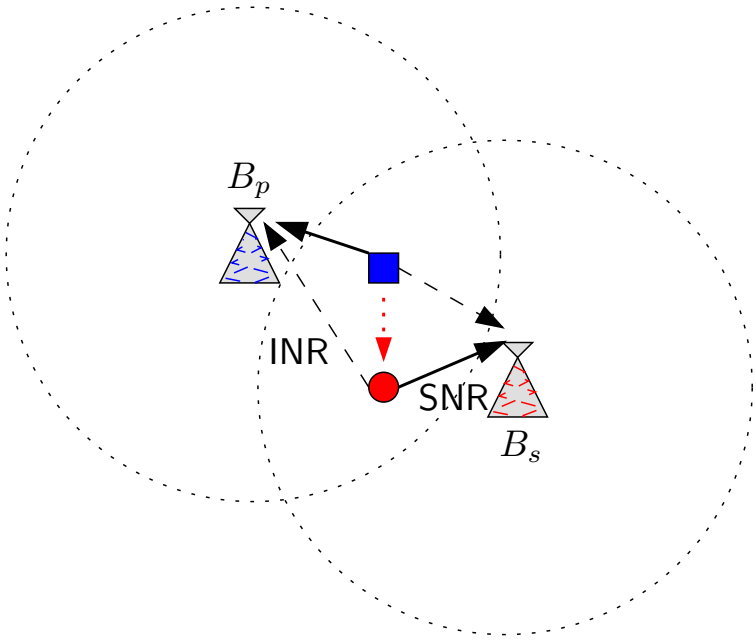

Figure 1: A possible arrangement of the primary and secondary receivers, base-stations B_p and B_s, respectively. The cognitive secondary user is represented by the circle and the primary user is represented by the square. The side-information path is depicted by the dotted line.

and jointly design their encoder-decoder pairs and then show that the optimal communication scheme for this cooperative situation has the property that the primary decoder does not depend on the encoder and decoder used by the secondary system. This cooperative communication scenario can be thought of as an interference channel [1], [16], [4] but with degraded message sets[2]: Achievable schemes for this channel have been first studied in [6]. A related problem of communicating a single private message along with a common message to each of the receivers has been studied in [12].

Furthermore, we exhibit a regime in which joint code design *is* beneficial when one considers the largest set of simultaneously achievable rates of the primary and cognitive users. We show that, unlike in the low-interference-gain regime, knowledge of the code used by the cognitive radio is required by the primary decoder in order to achieve all the rates in the capacity region of this interference channel when $\mathsf{INR} \gg \mathsf{SNR}$.

The rest of this paper is organized as follows. We first present the Gaussian *cognitive channel* in Section 2. We state our main result, the capacity of the cognitive channel in the low-interference-gain regime $\mathsf{INR} \leq \mathsf{SNR}$, in Section 3. The proof of our main result is given in Section 4, where we demonstrate the capacity region of the underlying interference channel with degraded message sets which inherently allows for joint code design. We then show that the benefit of joint code design becomes apparent in the high-interference-gain regime $\mathsf{INR} \gg \mathsf{SNR}$; this is done in Section 4.2.5. Finally, we study the

[2]The primary radio has only a subset of the messages available to the cognitive radio.

system-level implications of the optimal cognitive communication scheme in Section 5.

2 The Channel Model and Problem Statement

2.1 The cognitive channel

Consider the following communication scenario which we will refer to as the *cognitive channel*.

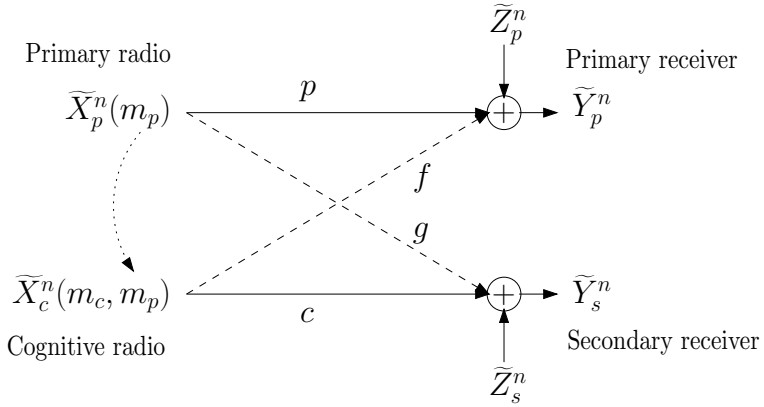

Figure 2: The (Gaussian) cognitive channel after n channel uses. The dashed lines represent interfering receptions. The dotted line represents the side-information path. The power constraints are $\widetilde{P}_p$ and $\widetilde{P}_c$ and noise variances are N_p and N_s.

The additive noise at the primary and secondary receivers, $\widetilde{Z}_p^n := (\widetilde{Z}_{p,1}, \widetilde{Z}_{p,2}, \ldots, \widetilde{Z}_{p,n})$ and $\widetilde{Z}_s^n := (\widetilde{Z}_{s,1}, \widetilde{Z}_{s,2}, \ldots, \widetilde{Z}_{s,n})$, is assumed to be i.i.d. across symbol times $i = 1, 2, \ldots n$ and distributed according to $\mathcal{N}(0, N_p)$ and $\mathcal{N}(0, N_s)$, respectively[3]. The correlation between $\widetilde{Z}_p^n$ and $\widetilde{Z}_s^n$ is irrelevant from the standpoint of probability of error or capacity calculations since the base-stations are not allowed to pool their signals. The primary user has message $m_p \in \{0, 1, \ldots, 2^{nR_p}\}$ intended for the primary receiver to decode, the cognitive user has message $m_c \in \{0, 1, \ldots, 2^{nR_c}\}$ intended for the secondary receiver *as well* as the message m_p of the primary user. The average power of the transmitted signals is constrained by $\widetilde{P}_p$ and $\widetilde{P}_c$, respectively:

$$\|\widetilde{X}_p^n\|^2 \leq n\widetilde{P}_p, \quad \|\widetilde{X}_c^n\|^2 \leq n\widetilde{P}_c. \tag{1}$$

The received signal-to-noise ratios (SNRs) of the desired signals at the primary and secondary base-station are $p^2\widetilde{P}_p/N_p$ and $c^2\widetilde{P}_c/N_s$, respectively. The received SNRs of

[3]Throughout the paper we will denote vectors in $\mathbb{R}^n$ by $X^n := (X_i, X_2, \ldots, X_n)$

the interfering signals at the primary and secondary base-station (INRs) are $f^2\widetilde{P}_c/N_p$ and $g^2\widetilde{P}_p/N_s$, respectively. The constants (p, c, f, g) are assumed to be real, positive and globally known. The results of this paper easily extend to the case of complex coefficients (see Section 5.3). The channel can be described by the pair of per-time-sample equations

$$\widetilde{Y}_p = p\widetilde{X}_p + f\widetilde{X}_c + \widetilde{Z}_p, \tag{2}$$
$$\widetilde{Y}_s = g\widetilde{X}_p + c\widetilde{X}_c + \widetilde{Z}_s, \tag{3}$$

where $\widetilde{Z}_p$ is $\mathcal{N}(0, N_p)$ and $\widetilde{Z}_s$ is $\mathcal{N}(0, N_s)$.

2.2 Transformation to standard form

We can convert every cognitive channel with gains (p, f, g, c), power constraints $(\widetilde{P}_p, \widetilde{P}_c)$ and noise powers (N_p, N_s) to a corresponding *standard form* cognitive channel with gains $(1, a, b, 1)$, power constraints (P_p, P_c) and noise powers $(1, 1)$, expressed by the pair of equations

$$Y_p = X_p + aX_c + Z_p, \tag{4}$$
$$Y_s = bX_p + X_c + Z_s, \tag{5}$$

where

$$a := \frac{f\sqrt{N_s}}{c\sqrt{N_p}}, \qquad b := \frac{g\sqrt{N_p}}{p\sqrt{N_s}},$$
$$P_p := \frac{p^2\widetilde{P}_p}{N_p}, \qquad P_c := \frac{c^2\widetilde{P}_c}{N_s}. \tag{6}$$

The capacity of this cognitive channel is the same as that of the original channel since the two channels are related by invertible transformations[4] that are given by

$$X_p := \frac{p\widetilde{X}_p}{\sqrt{N_p}}, \quad Y_p := \frac{\widetilde{Y}_p}{\sqrt{N_p}}, \quad Z_p := \frac{\widetilde{Z}_p}{\sqrt{N_p}}; \tag{7}$$
$$X_c := \frac{c\widetilde{X}_c}{\sqrt{N_s}}, \quad Y_s := \frac{\widetilde{Y}_s}{\sqrt{N_s}}, \quad Z_s := \frac{\widetilde{Z}_s}{\sqrt{N_s}}. \tag{8}$$

In deriving our main result we will consider this standard form of the cognitive channel without loss of generality and we will refer to it as the *cognitive* $(1, a, b, 1)$ *channel*.

[4]These transformations were used in [1], [3] and [16], in the context of the classical interference channel.

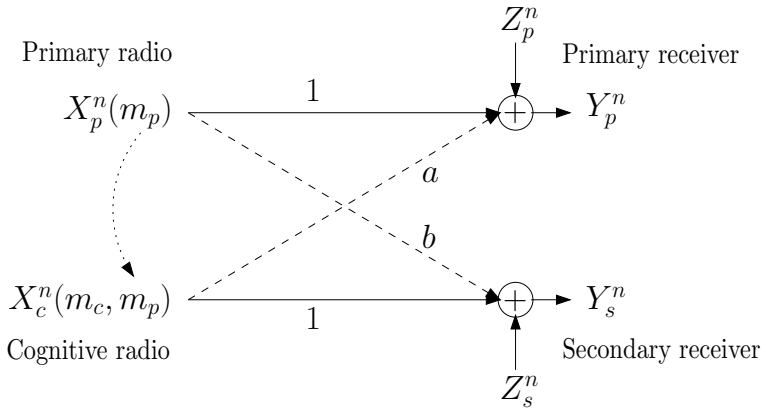

Figure 3: The cognitive channel in standard form. The channel gains (p, f, g, c) in the original channel are mapped to $(1, a, b, 1)$, powers $(\widetilde{P}_p, \widetilde{P}_c)$ are mapped to (P_p, P_c), and noise variances (N_p, N_s) are mapped to $(1, 1)$.

2.3 Coding on the cognitive channel

Let the channel input alphabets of the primary and cognitive radios be $\mathcal{X}_p = \mathbb{R}$ and $\mathcal{X}_c = \mathbb{R}$, respectively. Similarly, let the channel output alphabets at the primary and secondary receivers be $\mathcal{Y}_p = \mathbb{R}$ and $\mathcal{Y}_s = \mathbb{R}$, respectively.

The primary receiver is assumed to use a standard single-user decoder to decode $m_p \in \{1, 2, \ldots, 2^{nR_p}\}$ from Y_p^n, just as it would in the absence of the secondary system: Any decoder which achieves the AWGN channel capacity, such as the maximum-likelihood decoder or the joint-typicality decoder, will suffice. Following standard nomenclature, we say that R_p is *achievable* for the primary user if there exists a sequence (indexed by n) of encoding maps, $E_p^n : \{1, 2, \ldots, 2^{nR_p}\} \mapsto \mathcal{X}_p^n$, satisfying $\|X_p^n\|^2 \leq nP_p$, and for which the average probability of decoding error (average over the messages) vanishes as $n \to \infty$.

The cognitive radio is assumed to have knowledge of m_p, hence we have the following definition:

Definition 2.1 (Cognitive code) *A cognitive $(2^{nR_c}, n)$ code is a choice of an encoding rule (whose output we denote by X_c^n)*

$$E_c^n \;:\; \{1, 2, \ldots, 2^{nR_p}\} \times \{1, 2, \ldots, 2^{nR_c}\} \to \mathcal{X}_c^n, \tag{9}$$

such that $\|X_c^n\|^2 \leq nP_c$, and a choice of a decoding rule

$$D_c^n \;:\; \mathcal{Y}_s^n \to \{1, 2, \ldots, 2^{nR_c}\}. \tag{10}$$

The following key definition formalizes the important notion of *coexistence conditions* that the cognitive secondary system must satisfy.

Definition 2.2 (Achievability: cognitive user) *A rate R_c is said to be achievable for the cognitive user on a cognitive $(1, a, b, 1)$ channel if there exists a sequence of cognitive $(2^{nR_c}, n)$ codes such that the following two constraints are satisfied:*

1. *The average probability of error vanishes as $n \to \infty$, i.e.,*

$$P_{e,c}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}(D_c^n(Y_s^n) \neq j | m_p = i, m_c = j) \to 0; \tag{11}$$

2. *A rate of $R_p^* \stackrel{\text{def}}{=} \frac{1}{2}\log(1 + P_p)$ is achievable for the primary user.*

Definition 2.3 (Capacity) *The* capacity *of the cognitive channel is defined to be the largest achievable rate R_c for the cognitive user.*

Our main result, presented in the following section, precisely quantifies the capacity of the cognitive channel in the "low-interference-gain" regime.

3 The Main Result

If the received SNR of the cognitive radio transmission is lesser at the primary receiver than at the secondary receiver, we say that the primary system is affected by a *low interference gain*. This is the case that is most likely to occur in practice since the cognitive radio is typically closer to its intended receiver (the secondary base-station) than to the primary base-station. In terms of the parameters of our problem, this situation corresponds to $f\sqrt{N_s} \leq c\sqrt{N_p}$ in our original cognitive channel, or, equivalently, to $a \leq 1$ in the corresponding standard-form cognitive $(1, a, b, 1)$ channel. Our main result is an explicit expression for the capacity of the cognitive channel in this regime.

Theorem 3.1 *The capacity of the cognitive $(1, a, b, 1)$ channel is*

$$R_c^* = \frac{1}{2}\log(1 + (1 - \alpha^*)P_c), \tag{12}$$

as long as $a \leq 1$. The constant $\alpha^* \in [0, 1]$ is defined in (17).

Note that Theorem 3.1 holds for any $b \in \mathbb{R}$ (or equivalently any $p, g \in \mathbb{R}$ in the original cognitive channel).

4 Proof of the Main Result

4.1 The forward part

To show the existence of a capacity-achieving cognitive $(2^{nR_c^*}, n)$ code, we generate a sequence of random codes such that the average probability of error (averaged over the ensemble of codes and messages) vanishes as $n \to \infty$. In particular, we have the following codes:

- E_p^n *ensemble:* Given $m_p \in \{1, 2, \ldots, 2^{nR_p}\}$, generate the codeword $X_p^n \in \mathbb{R}^n$ by drawing its coordinates i.i.d. according to $\mathcal{N}(0, P_p)$.
- E_c^n *ensemble:* Since the cognitive radio knows m_p as well as E_p^n, it can form X_p^n and perform superposition coding as follows:

$$X_c^n = \hat{X}_c^n + \sqrt{\frac{\alpha P_c}{P_p}} X_p^n, \tag{13}$$

  where $\alpha \in [0, 1]$. The codeword $\hat{X}_c^n$ encodes $m_c \in \{1, 2, \ldots, 2^{nR_c}\}$ and is generated by performing *Costa precoding* [3] (also known as *dirty-paper coding*) treating $(b + \sqrt{\alpha \frac{P_c}{P_p}}) X_p^n$ as non-causally known interference that will affect the secondary receiver in the presence of $\mathcal{N}(0, 1)$ noise. The encoding is done by *random binning* [3].
- D_c^n: Costa decoder (having knowledge of the binning encoder E_c^n) [3].

The key result of Costa [4] is that, using the dirty-paper coding technique, the maximum achievable rate is the same as if the interference was also known at the receiver, i.e., as if it were absent altogether. The characteristic feature of this scheme is that the resulting codeword $\hat{X}_c^n$ is statistically independent of X_p^n and is i.i.d. Gaussian. To satisfy the average power constraint of P_c on the components of X_c^n, each coordinate of $\hat{X}_c^n$ must, in fact, be $\mathcal{N}(0, (1 - \alpha)P_c)$. Hence, the primary receiver can treat $\hat{X}_c^n$ as independent Gaussian noise. Using standard methodology, it can be shown that the average probability of error for decoding m_p (averaged over the code ensembles and messages) vanishes, as $n \to \infty$, for all rates R_p below

$$\frac{1}{2} \log \left(1 + \frac{(\sqrt{P_p} + a\sqrt{\alpha P_c})^2}{1 + a^2(1 - \alpha)P_c} \right). \tag{14}$$

Similarly, the average probability of error in decoding m_c vanishes for all rates R_c below

$$\frac{1}{2} \log(1 + (1 - \alpha)P_c). \tag{15}$$

However, in order to ensure that a given rate is *achievable* for the cognitive user in the sense of Definition 2.2, we must have that

$$\frac{1}{2}\log\left(1+\frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right)=\frac{1}{2}\log(1+P_p)=:R_p^*. \tag{16}$$

Observe that, if $a=0$, any choice of $\alpha\in[0,1]$ will satisfy (16): in this case we should set $\alpha^*=0$ to maximize the rate achievable for the cognitive user. For $0<a\leq 1$, by the Intermediate Value Theorem, this quadratic equation in α always has a unique root in $[0,1]$:

$$\alpha^*=\left(\frac{\sqrt{P_p}\left(\sqrt{1+a^2P_c(1+P_p)}-1\right)}{a\sqrt{P_c}(1+P_p)}\right)^{\frac{1}{2}}. \tag{17}$$

Finally, since the code-ensemble-averaged (and message-averaged) probabilities of error vanish, there must exist a particular sequence of cognitive codes and primary encoders for which the (message-averaged) probabilities of error vanish as well. Hence, $R_c^*=\frac{1}{2}\log(1+(1-\alpha^*)P_c)$ is achievable for the cognitive user in the sense of Definition 2.2.

4.2 The converse part

4.2.1 Proof outline

In order to prove the converse to our main result we will first relax the constraints of our problem and allow for *joint* primary and cognitive code design. This relaxation leads naturally to an interference channel with degraded message sets[5], which we will abbreviate as IC-DMS for convenience.

Our approach is to first characterize the capacity region of the IC-DMS, i.e., the largest set of rate tuples (R_p,R_c) at which joint reliable communication can take place. We then make the key observation that the joint coding scheme that achieves all the rate tuples in the capacity region of the IC-DMS has the property that the decoder at the primary receiver is a standard single-user decoder. Furthermore, we show that there exists a point $(R_p,R_c)=(R_p^*,R_c^*)$ on the boundary of the capacity region of the IC-DMS, where $R_p^*=\frac{1}{2}\log(1+P_p)$ and $R_c^*=\frac{1}{2}\log(1+(1-\alpha^*)P_c)$ with α^* given by (17). We then conclude that $R_c=R_c^*$ is the capacity of the corresponding cognitive channel.

[5]The primary user knows m_p while the cognitive user knows $\{m_p,m_c\}$, hence the primary user has a subset of the messages available to the cognitive user.

4.2.2 Joint code design: The IC-DMS

The input-output equations of the IC-DMS, as for the cognitive channel, are given by (2), (3) with the standard form given by (4), (5). We will denote the IC-DMS in standard form by "$(1, a, b, 1)$-IC-DMS".

Definition 4.1 (IC-DMS code) *A $(2^{nR_p}, 2^{nR_c}, n)$ code for the $(1, a, b, 1)$-IC-DMS is a choice of an encoding rule and a decoding rule: The encoding rule is a pair of maps (whose outputs we denote by X_p^n and X_c^n, respectively)*

$$e_p^n \;:\; \{1, 2, \ldots, 2^{nR_p}\} \to \mathcal{X}_p^n, \tag{18}$$
$$e_c^n \;:\; \{1, 2, \ldots, 2^{nR_p}\} \times \{1, 2, \ldots, 2^{nR_c}\} \to \mathcal{X}_c^n, \tag{19}$$

such that $\|X_p^n\|^2 \leq nP_p$ and $\|X_c^n\|^2 \leq nP_c$. The decoding rule is a pair of maps

$$d_p^n \;:\; \mathcal{Y}_p^n \to \{1, 2, \ldots, 2^{nR_p}\}, \tag{20}$$
$$d_c^n \;:\; \mathcal{Y}_s^n \to \{1, 2, \ldots, 2^{nR_c}\}. \tag{21}$$

Given that the messages selected are $(m_p = i, m_c = j)$, an error occurs if $d_p^n(Y_p^n) \neq i$ or $d_c^n(Y_s^n) \neq j$.

Definition 4.2 (Achievability: IC-DMS) *A rate vector (R_p, R_c) is said to be* achievable *if there exists a sequence of $(2^{nR_p}, 2^{nR_c}, n)$ codes such that the average probability of error at each of the receivers vanishes as $n \to \infty$, i.e.,*

$$\widetilde{P}_{e,p}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}(d_p^n(Y_p^n) \neq i | m_p = i, m_c = j) \to 0, \tag{22}$$

$$P_{e,s}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}(d_c^n(Y_s^n) \neq j | m_p = i, m_c = j) \to 0. \tag{23}$$

Definition 4.3 (Capacity region) *The capacity region of the IC-DMS is the closure of the set of achievable rate vectors (R_p, R_c).*

4.2.3 The capacity region of the IC-DMS under a low interference gain

The following theorem characterizes the capacity region of the $(1, a, b, 1)$-IC-DMS with $a \leq 1$ and arbitrary $b \in \mathbb{R}$.

Theorem 4.1 *The capacity region of the* $(1, a, b, 1)$*-IC-DMS with* $a \leq 1$ *and* $b \in \mathbb{R}$ *is given by the union, over all* $\alpha \in [0, 1]$*, of the rate regions*

$$0 \leq R_p \leq \frac{1}{2} \log \left(1 + \frac{(\sqrt{P_p} + a\sqrt{\alpha P_c})^2}{1 + a^2(1-\alpha)P_c} \right), \tag{24}$$

$$0 \leq R_c \leq \frac{1}{2} \log \left(1 + (1-\alpha)P_c \right). \tag{25}$$

Proof of achievability: The random coding scheme described in the forward part of the proof of Theorem 3.1 (Section 4.1) achieves the rates (24) and (25) stated in the theorem. We emphasize that, in this scheme, the primary receiver employs a single-user decoder.

Proof of converse: See Appendix A.

4.2.4 The capacity of the cognitive channel under a low interference gain

The proof of Theorem 4.1 reveals that the jointly designed code that achieves all the points on the boundary of the capacity region of the IC-DMS is such that the primary receiver uses a standard single-user decoder, just as it would *in the absence* of the cognitive radio. In other words, the primary decoder d_p^n does not depend on e_c^n and d_c^n. Thus, in order to find the largest rate that is achievable by the cognitive user in the sense of Definition 2.2 we can without loss of generality restrict our search to the boundary of the capacity region of the underlying IC-DMS. Hence, to find this capacity of the cognitive channel, we must solve for the positive root of the quadratic equation (16) in α. The solution is given by α^* in (17), hence the capacity is

$$R_c^* = \frac{1}{2} \log(1 + (1 - \alpha^*)P_c). \tag{26}$$

Thus we have established the proof of Theorem 3.1. □

The proof of the converse of Theorem 4.1 allows us to characterize the sum-capacity of the $(1, a, b, 1)$-IC-DMS for any $a \geq 1$ and the entire capacity region if a is sufficiently large. These two ancillary results are shown in the following section.

4.2.5 The high-interference-gain regime

The sum-capacity for $a \geq 1$

Corollary 4.1 *The maximum of $R_p + R_c$ over all (R_p, R_c) in the capacity region of the $(1, a, b, 1)$-IC-DMS with $a \geq 1$ and $b \in \mathbb{R}$ is achieved with $\alpha = 1$ in (24) and (25), i.e.,*

$$C_{sum}(a) = \frac{1}{2}\log\left(1 + \left(\sqrt{P_p} + a\sqrt{P_c}\right)^2\right). \tag{27}$$

Proof: See Appendix B

Contrary to the development so far, in the following section we will observe that, in the very-high-interference-gain regime, the optimal (jointly designed) IC-DMS code is such that the primary decoder d_p^n depends on the cognitive encoder e_c^n.

The benefit of joint code design

When the interference gain at the primary receiver due to the cognitive radio transmissions (parameter a) is sufficiently large, the optimal decoder at the primary receiver of the IC-DMS is one that decodes the message of the cognitive user before decoding the message of the primary user.

First, we demonstrate an achievable scheme in the following lemma.

Lemma 4.2 *Consider the cognitive $(1, a, b, 1)$-interference channel. For every $\alpha \in [0, 1]$, the rate pair (R_p, R_c) satisfying*

$$R_p = \hat{R}_p(\alpha) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \left(\sqrt{P_p} + a\sqrt{\alpha P_c}\right)^2\right), \tag{28}$$

$$R_c = \hat{R}_c(\alpha) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \frac{(1-\alpha)P_c}{1 + (b\sqrt{P_p} + \sqrt{\alpha P_c})^2}\right), \tag{29}$$

is achievable as long as

$$a \geq \frac{\sqrt{\alpha P_p P_c}}{K(\alpha)} + \sqrt{K(\alpha) + P_p\left(1 + (b\sqrt{P_p} + \sqrt{\alpha P_c})^2\right)}, \tag{30}$$

where $K(\alpha) \stackrel{\text{def}}{=} 1 + b^2 P_p + 2b\sqrt{\alpha P_p P_c}$.

Proof: The primary transmitter forms X_p^n by drawing its coordinates i.i.d. according to $\mathcal{N}(0, P_p)$. Since the cognitive radio knows m_p and e_p^n it forms X_p^n then generates X_c^n by superposition coding:

$$X_c^n = \hat{X}_c^n + \sqrt{\frac{\alpha P_c}{P_p}} X_p^n,$$

where $\hat{X}_c^n$ is formed by drawing its coordinates i.i.d. according to $\mathcal{N}(0, \sqrt{(1-\alpha)P_c})$ for some $\alpha \in [0,1]$. The decoder d_p^n at the primary receiver first decodes m_c treating $(1 + a\sqrt{\alpha P_c/P_p})X_p^n$ as independent Gaussian noise. It then reconstructs $a\hat{X}_c^n$ (which it can do because it knows e_c^n) and subtracts off its contribution from Y_p^n before decoding m_p. The decoding rule d_c^n at the secondary receiver is simply to decode m_c treating $(b+\sqrt{\alpha P_c/P_p})X_p^n$ as independent Gaussian noise. The rates achievable with this scheme are then exactly given by (28) and (29), provided that the rate at which the primary receiver can decode the cognitive user's message is not the limiting factor, i.e.,

$$\frac{(1-\alpha)P_c}{1+(b\sqrt{P_p}+\sqrt{\alpha P_c})^2} \leq \frac{a^2(1-\alpha)P_c}{1+\left(\sqrt{P_p}+a\sqrt{\alpha P_c}\right)^2}.$$

Solving this quadratic inequality for a, we find that the condition is satisfied only when a satisfies inequality (30) stated in the theorem. □

Theorem 4.3 *A point (R_p, R_c) is on the boundary of the capacity region of the cognitive $(1, a, b, 1)$-interference channel if there exists $\alpha \in [0,1]$ such that*

1. *$(R_p, R_c) = (\hat{R}_p(\alpha), \hat{R}_c(\alpha))$ where $\hat{R}_p(\alpha)$ and $\hat{R}_c(\alpha)$ are defined in (28) and (29), respectively,*
2. *a and b satisfy the condition given in (30), and*
3. *$b \leq b_{\max}(\mu_\alpha, a)$ where $\mu_\alpha \stackrel{\text{def}}{=} -\left.\frac{d^- \hat{R}_c(x)}{d\hat{R}_p(x)}\right|_{x=\alpha}$ and $b_{\max}(\mu, a)$ is defined in Appendix C.*

Proof of achievability: Given in Lemma 4.2.

Proof of converse: Given in Appendix C.

Observe that Theorem 4.3 characterizes the entire capacity region of the $(1, a, b, 1)$-IC-DMS with $a \geq \sqrt{P_p P_c}/K(1) + \sqrt{K(1) + P_p\left(1 + (b\sqrt{P_p}+\sqrt{P_c})^2\right)}$ and $b \leq b_{\max}(\mu_\alpha, a)$.

5 System-level Considerations

In this section we use our results on the capacity-achieving cognitive communication scheme to derive insight into a practical implementation of cognitive radios.

5.1 Properties of the optimal scheme

5.1.1 Avoiding the "hidden-terminal" problem

The network of Fig. 1 models the situation in which the geographic location of B_s is not assigned in accordance with any centralized cell-planning policy and it can be arbitrarily close to B_p. Consequently, the secondary users that are in close proximity to B_p could potentially cause significant interference for the primary system if the secondary system is to operate over the same frequency band.

One possible adaptive communication scheme that the cognitive radio could employ in order to avoid interfering with the primary user in its vicinity would be to restrict its transmissions to only the time-frequency slots which are not occupied by the signals of the detected primary radio. Indeed, this idea of "opportunistic" orthogonal communication was what led to the birth of the notion of cognitive radio. However, one drawback of such a protocol is that the cognitive radio would very likely cause interference to other, more distant, primary users whose presence – i.e., time-frequency locations – it could not detect. The degradation in overall performance of the primary system due to this "hidden-terminal" problem could potentially be significant[6], especially in the context of OFDMA [9], [10] where the primary users are allocated orthogonal time-frequency slots and the SINR required for decoding is typically large.

Contrary to this, we find that the optimal strategy is for the cognitive radio to simultaneously transmit in the same frequency slot as that used by the primary user in its vicinity. An immediate benefit of this scheme is that, if the transmissions of different primary users are mutually orthogonal, the cognitive radio can *only* (potentially) affect the performance achievable by the primary radio whose codeword it has decoded. Furthermore, we know that a proper tuning of the parameter α can, in fact, ensure that the primary user's rate is unaffected.

5.1.2 Robustness to noise statistics

All our results have been derived under the assumption that the noise affecting the receivers, Z_p^n and Z_s^n, is i.i.d. Gaussian. In [2] it was shown that using a Costa encoder-decoder pair that is designed for additive i.i.d Gaussian noise on a channel with arbitrary (additive) noise statistics will cause no loss in the achievable rates.[7] Combined with the similar classical result for the standard AWGN channel [11], we see that the maximal rate expressed in Theorem 3.1 is achievable for all noise distributions.

[6]Classical RTS/CTS solutions to this problem are not viable since they require that the primary system *ask* for access to the very spectrum that it owns.

[7]Note that this is an achievability result: The capacity of the channel with this arbitrary noise could be larger but a different code would be required to achieve it.

5.2 Obtaining the side-information

In practice, the cognitive radio must obtain the primary radio's codeword in a causal fashion – its acquisition thus introducing delays in the cognitive radio transmissions[8]. In a typical situation, due to its relative proximity to the primary user, the cognitive radio can receive the primary transmissions with a greater received SNR than that experienced by the primary receiver. Hence, it seems plausible that the cognitive radio could decode[9] the message of the primary user in fewer channel uses than are required by the primary receiver. Recent work in distributed space-time code design [13] indicates that this overhead decoding delay is negligible if the cognitive radio has as little as a 10 dB advantage in the received SNR over the primary receiver.

5.3 Extension to complex baseband

The results of this paper can easily be extended to the case in which the channel gains are complex quantities, i.e., $p, f, g, c \in \mathbb{C}$ in the case of the original cognitive (p, f, g, c) channel with power constraints (P_p, P_c) and noise variances (N_p, N_s), as defined in Section 2.1. However, the optimal cognitive encoder rule (13) must change slightly: The superposition scheme takes the form

$$X_c^n = \hat{X}_c^n + \frac{f^*}{|f|} e^{j\theta_p} \sqrt{\alpha \frac{P_c}{P_p}} X_p^n, \tag{31}$$

where $p = |p|e^{j\theta_p}$. The codeword $\hat{X}_c^n$ is again generated by Costa precoding, but the assumed interference at the secondary receiver is now

$$\left(\frac{g}{c} + \frac{f^*}{|f|} e^{j\theta_p} \sqrt{\alpha \frac{P_c}{P_p}} \right) X_p^n, \tag{32}$$

and the assumed noise is $\mathcal{CN}(0, N_s/|c|^2)$. The factor $e^{j\theta_p}$ in (31) essentially implements transmit beamforming to the primary receiver, hence ensuring that all the rates given by

$$0 \le R_p \le \log \left(1 + \frac{\left(|p|\sqrt{P_p} + |f|\sqrt{\alpha P_c} \right)^2}{N_p + |f|^2 (1-\alpha) P_c} \right), \tag{33}$$

$$0 \le R_c \le \log \left(1 + \frac{|c|^2 (1-\alpha) P_c}{N_s} \right), \tag{34}$$

are achieved in the underlying IC-DMS. As before, we can then choose $\alpha = \alpha^*$ (determined by (17)), so that $R_c^* = \log(1 + |c|^2(1-\alpha^*)P_c/N_s)$ is achievable in the spirit of Definition 2.2 but with $R_p^* = \log(1 + |p|^2 P_p/N_p)$.

[8] Under a half-duplex constraint the cognitive radio must first "listen" in order to decode the primary message before it can use this side-information for its own transmission.

[9] The cognitive radio is assumed to know the encoder of the primary user.

5.4 Communicating without channel-state feedback from the primary base-station

In order to perform the complex base-band superposition coding scheme (31) and, implicitly, the Costa precoding for known interference (32), the cognitive radio must know each of the four parameters g, c, f and p, both in magnitude and phase. To obtain estimates for p and f, the cognitive radio would require feedback from the primary base-station. In section Section 5.5, we discuss ways in which the estimation and feedback of these parameters could be implemented. In this section, however, we present an alternative (suboptimal) scheme which requires no feedback from the primary base-station.

Suppose that, after having decoded X_p^n, the cognitive radio transmits the following n-symbol codeword:

$$X_c^n = \hat{X}_c^n + \sqrt{\alpha \frac{P_c}{P_p}} X_p^n, \tag{35}$$

where the codeword $\hat{X}_c^n$ is generated by Costa precoding for the interference

$$\left(\frac{g}{c} + \sqrt{\alpha \frac{P_c}{P_p}} \right) X_p^n, \tag{36}$$

assuming the presence of $\mathcal{CN}(0, N_s/|c|^2)$ noise at the secondary base-station.

- *Obtaining c:* The parameter c could be estimated at the secondary base-station by using the cognitive radio's pilot signal or in a decision-directed fashion. The estimate could then be fed back to the cognitive radio.
- *Obtaining g:* If the secondary base-station synchronizes to the primary radio's pilot signal, it could estimate g during the time the cognitive radio is in its silent "listening" phase and then feed this estimate back to the cognitive radio. Alternatively, if the cognitive radio reveals to the secondary base-station the code used by the primary radio, the secondary base-station could use the silent "listening" phase to decode a few symbols transmitted by the primary radio thereby estimating the parameter g.

We can express the received discrete-time base-band signal at the primary base-station at time sample m as

$$Y_p[m] \;=\; pX_p[m] + f\sqrt{\alpha \frac{P_c}{P_p}} X_p[m - l_c] + Z_{\text{total}}[m], \tag{37}$$

where $Z_{\text{total}}[m] = f\hat{X}_c[m - l_c] + Z_p[m]$ is the aggregate noise. The integer l_c accounts for the delay incurred while the cognitive radio "listens" and decodes the primary codeword

before it transmits its own signal. This equation essentially describes a time-invariant two-tap ISI channel for the primary transmission, hence we can apply a Rake receiver (in the case the primary system uses direct-sequence spread-spectrum) or transmit-receive architectures such as OFDM[10] to extract both a diversity gain of two and a power gain of $|p|^2\widetilde{P}_p + |f|^2\alpha P_c$ at the primary base-station (see, for instance, Chapter 3 of [18], and references therein). Given $\alpha \in [0,1]$, the rates achievable by the primary and cognitive users using such a scheme are given by

$$0 \leq R_p \leq \log\left(1 + \frac{|p|^2 P_p + |f|^2 \alpha P_c}{N_p + |f|^2(1-\alpha)P_c}\right), \tag{38}$$

$$0 \leq R_c \leq \log\left(1 + \frac{|c|^2(1-\alpha)P_c}{N_s}\right). \tag{39}$$

In order to avoid causing interference to the primary user, the following equation must be satisfied:

$$\frac{|p|^2 P_p + |f|^2 \alpha P_c}{N_p + |f|^2(1-\alpha)P_c} = \frac{|p|^2 P_p}{N_p}, \tag{40}$$

If the cognitive radio tunes its parameter α such that

$$\alpha = \frac{|p|^2 P_p/N_p}{1 + |p|^2 P_p/N_p}, \tag{41}$$

this condition will be satisfied, hence $R_p = R_p^*$. Expression (41) confirms the intuitive notion that, if the primary system is operating at high SNR, the cognitive radio should not interfere with it, i.e., α should be close to one.

From (41), we see that, in order to design the optimal α, the cognitive radio only needs to know the received SNR of the primary transmission at the primary base-station: $|p|^2 P_p/N_p$. If the primary system uses a good (capacity-achieving) AWGN channel code and the cognitive radio knows this, the cognitive radio can easily compute an estimate of this received SNR since it knows the rate at which the primary user is communicating, R_p: This estimate is simply given by $e^{R_p} - 1$. Thus, an immediate benefit of this scheme is that the primary base-station need not feed-back the parameters f and p at all: The cognitive radio can perform completely autonomously.

Though expression (41) does not depend on $|f|$, we can see that (40) can approximately be satisfied even with $\alpha = 0$ when $|f|^2$ is very small. Since the cognitive radio has no information about $|f|$ and, in practice, may not even be able to obtain $|p|^2 P_p/N_p$ (if the primary system is not using a good AWGN code), a natural way for the cognitive radio to enter the spectrum of the primary would be by slowly *ramping* up its power

[10]The primary base-station would most likely already employ one of these schemes as a means of dealing with the multi-path point-to-point channel between the primary radio and itself. In the context of OFDM, however, the cyclic prefix would have to be long enough to account for the extra delay-spread introduced by the cognitive radio's transmission.

P_c from 0 and decreasing α from 1 while simultaneously listening for the Automatic Repeat Request (ARQ) control signal from the primary base-station. Once this signal is detected, the cognitive radio would either slightly decrease P_c or increase α until the primary base-station stops transmitting ARQs[11].

5.5 Obtaining the channel-state information

In order to implement the optimal communication scheme of Costa coding and beamforming (31), the cognitive radio must obtain estimates of p and f from the primary base-station. We present the following simple algorithm for estimation and feedback of these parameters:

1. At first, the cognitive user is silent and the primary base-station broadcasts the current estimate of p, call it $\hat{p}$, along with the primary user's ID on the control channel to which the cognitive radio is tuned. The primary base-station is assumed to be able to track p by either using a pilot signal or in a decision-directed fashion. Thus, the cognitive radio can obtain $\hat{p}$.

2. Upon entering the system and decoding the message of the primary user in its vicinity, the cognitive radio simply performs amplify-and-forward relaying of the primary codeword:

$$X_c^n = \sqrt{\frac{P_c}{P_p}} X_p^n. \tag{42}$$

3. The primary base-station receives

$$\left(p + f\sqrt{\frac{P_c}{P_p}}\right) X_p^n + Z_p^n, \tag{43}$$

   hence it can compute an estimate, $\hat{h}$, of the overall channel gain $\left(p + f\sqrt{\frac{P_c}{P_p}}\right)$ as it decodes m_p.

4. The quantized version of $\hat{h}$ is then broadcast on the control channel along with the given primary user's ID.

5. The cognitive radio picks up this information from the control channel and then computes $\hat{h} - \hat{p}$.

6. The quantity $\hat{h} - \hat{p}$ is an estimate for $f\sqrt{P_c/P_p}$ which is then multiplied by $\sqrt{P_p/P_c}$, to obtain an estimate for f.

[11] This scheme is analogous to the power control mechanism used in CDMA systems.

Note that it is possible that $\left|p + f\sqrt{P_c/P_p}\right| < |p|$ in step 3 above. In this case the primary system would momentarily not be able to support the requested rate of $\log(1+|p|^2 P_p/N_p)$ and an Automatic Repeat Request (ARQ) would be generated by the primary base-station. However, by this time, the cognitive radio would have already obtained the estimate of f and the next (repeated) transmission would be guaranteed to be successful.

A Proof of the converse part of Theorem 4.1

First we observe that the rate-region specified in Theorem 4.1 is a convex set in Proposition D.1. We will use the following standard result from convex analysis (see, for instance, [15]) in the proof of the converse.

Proposition A.1 *A point $\mathbf{R}^* = (R_p^*, R_c^*)$ is on the boundary of the a capacity region if and only if there exists a $\mu \geq 0$ such that the linear functional $\mu R_p + R_c$ achieves its maximum, over all (R_p, R_c) in the region, at $\mathbf{R}^*$.*

A.1 The $\mu \leq 1$ case

For convenience, we will consider a channel whose output at the primary receiver is normalized by a, i.e., a channel whose input-output single-letter equations are given by

$$\hat{Y}_p^n \stackrel{\text{def}}{=} \frac{1}{a}X_p^n + X_c^n + \frac{1}{a}Z_p^n, \tag{44}$$

$$Y_s^n = bX_p^n + X_c^n + Z_s^n. \tag{45}$$

Note that the capacity region of this channel is the same as that of the original channel (4), (5) since normalization is an invertible transformation.

Suppose that there exists a sequence of $(2^{nR_p}, 2^{nR_c}, n)$ codes for the IC-DMS (defined by the decoder maps (20), (21) and encoder maps (18), (19)) such that the average probability of error at both receivers vanishes as $n \to \infty$, i.e., $\widetilde{P}_{e,p}^{(n)} \to 0$, $P_{e,s}^{(n)} \to 0$, where the assumption is that the messages (m_p, m_c) are chosen uniformly and independently. Then, by Fano's inequality, $H(m_p|Y_p^n) \leq n\epsilon_{p,n}$ and $H(m_c|Y_s^n) \leq n\epsilon_{s,n}$, where $\epsilon_{p,n} \to 0$ and $\epsilon_{s,n} \to 0$ as $\widetilde{P}_{e,p}^{(n)} \to 0$, $P_{e,s}^{(n)} \to 0$, respectively. We start with the following bound on nR_p:

$$\begin{aligned} nR_p &\stackrel{(a)}{=} H(m_p), \\ &= I(m_p; \hat{Y}_p^n) + H(m_p|\hat{Y}_p^n), \\ &\stackrel{(b)}{\leq} I(m_p; \hat{Y}_p^n) + n\epsilon_{p,n}, \\ &= h(\hat{Y}_p^n) - h(\hat{Y}_p^n|m_p) + n\epsilon_{p,n}, \end{aligned} \tag{46}$$

where (a) follows since m_p and m_c are uniformly distributed on $\{1, 2, \ldots, 2^{nR_p}\}$ and $\{1, 2, \ldots, 2^{nR_p}\}$ respectively, (b) follows from Fano's inequality. Also, we have that,

$$\begin{aligned}
nR_c &= H(m_c), \\
&= H(m_c) + H(m_c|Y_s^n, m_p) - H(m_c|Y_s^n, m_p), \\
&= I(m_c; Y_s^n|m_p) + H(m_c|Y_s^n, m_p), \\
&\overset{(a)}{\leq} I(m_c; Y_s^n|m_p) + n\epsilon_{s,n}, \\
&= h(Y_s^n|m_p) - h(Y_s^n|m_p, m_c) + n\epsilon_{s,n}, \\
&\overset{(b)}{\leq} h(Y_s^n|m_p) - h(Y_s^n|m_p, m_c, X_p^n, X_c^n) + n\epsilon_{s,n}, \\
&\overset{(c)}{=} h(Y_s^n|m_p) - h(Z_s^n) + n\epsilon_{s,n}, \qquad (47)
\end{aligned}$$

where (a) follows from Fano's inequality and the fact that conditioning does not increase entropy, (b) follows from the fact that conditioning does not increase entropy, and (c) follows from the the fact that Z_s^n is independent of (m_p, m_c) and hence also of (X_p^n, X_c^n).

Let $\widetilde{Z}^n$ be a zero mean Gaussian random vector, independent of $(X_p^n, X_c^n, Z_p^n, Z_s^n)$ and with covariance matrix $(\frac{1}{a^2} - 1)\mathbf{I}_n$. Then, we can write

$$\begin{aligned}
h(\hat{Y}_p^n|m_p) &\overset{(a)}{=} h(\hat{Y}_p^n|m_p, X_p^n), \\
&\overset{(b)}{=} h\left(\hat{Y}_p^n - \frac{1}{a}X_p^n|m_p, X_p^n\right), \\
&= h\left(X_c^n + \frac{1}{a}Z_p^n|m_p, X_p^n\right), \\
&\overset{(c)}{=} h(X_c^n + Z_s^n + \widetilde{Z}^n|m_p, X_p^n), \\
&\overset{(d)}{=} h(X_c^n + Z_s^n + \widetilde{Z}^n|m_p), \\
&\overset{(e)}{=} h(\widetilde{Y}^n + \widetilde{Z}^n|m_p), \qquad (48)
\end{aligned}$$

where (a) and (d) hold since X_p^n is the output of a deterministic function of m_p, (b) holds because translation does not affect entropy, (c) follows from the fact that Gaussian distributions are infinitely divisible and from the definition of $\widetilde{Z}^n$ and (e) follows from the definition $\widetilde{Y}^n \overset{\text{def}}{=} X_c^n + Z_s^n$. By similar reasoning, we can write

$$h(Y_s^n|m_p) = h(\widetilde{Y}^n|m_p). \qquad (49)$$

Combining the bounds in (46) and (47), we get

$$\begin{aligned}
n(\mu R_p + R_c) &\leq \mu(h(\hat{Y}_p^n) - h(\hat{Y}_p^n|m_p)) + h(Y_s^n|m_p) - h(Z_s^n) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, \\
&\overset{(a)}{=} \mu h(\hat{Y}_p^n) + h(Y_s^n|m_p) - \mu h(\hat{Y}_p^n|m_p) - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, \\
&\overset{(b)}{=} \mu h(\hat{Y}_p^n) + h(\widetilde{Y}^n|m_p) - \mu h(\widetilde{Y}^n + \widetilde{Z}^n|m_p) - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, \\
&\overset{(c)}{\leq} \mu h(\hat{Y}_p^n) + h(\widetilde{Y}^n|m_p) - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}h(\widetilde{Y}^n|m_p)} + e^{\frac{2}{n}h(\widetilde{Z}^n)}\right) \\
&\qquad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, \quad (50)
\end{aligned}$$

where (a) follows from the fact that $Z_s^n \sim \mathcal{N}(0, \mathbf{I}_n)$, (b) follows from equalities (48) and (49), (c) follows from the conditional version of the Entropy Power Inequality (see Proposition D.2).

Let X_1^{j-1} denote the first $j-1$ components of the vector X^n with the understanding that X_1^0 is defined to be some constant and let X_j denote the j-th component. We can upper-bound $h(\widetilde{Y}^n|m_p)$ as follows:

$$\begin{aligned}
h(\widetilde{Y}^n|m_p) &= h(\widetilde{Y}^n|m_p, X_p^n), \\
&\overset{(a)}{=} \sum_{j=1}^{n} h(\widetilde{Y}_j|m_p, \widetilde{Y}_1^{j-1}, X_{p,j}, X_{p,1}^{j-1}), \\
&\overset{(b)}{\leq} \sum_{j=1}^{n} h(\widetilde{Y}_j|X_{p,j}), \\
&\overset{(c)}{\leq} \sum_{j=1}^{n} \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}[\widetilde{Y}_j^2] - \frac{\mathbb{E}[\widetilde{Y}_j X_{p,j}]^2}{\mathbb{E}[X_{p,j}^2]}\right)\right), \\
&\overset{(d)}{=} \sum_{j=1}^{n} \frac{1}{2}\log\left(2\pi e\left((1-\alpha_j)P_{c,j} + 1\right)\right), \quad (51) \\
&\overset{(e)}{\leq} \frac{n}{2}\log\left(2\pi e\left((1-\alpha)P_c + 1\right)\right), \quad (52)
\end{aligned}$$

where (a) follows from the chain rule and (b) follows from the fact that conditioning does not increase entropy, and (c) follows from Lemma D.1. Equality (d) follows from the following argument: Since jointly Gaussian $X_{p,j}$, $Y_{p,j}$ achieve equality in (c) (by Lemma D.1), we can without loss of generality, let

$$X_{c,j} = \hat{X}_{c,j} + \sqrt{\alpha_j \frac{P_{c,j}}{P_{p,j}}} X_{p,j}, \quad (53)$$

where $\hat{X}_{c,j} \sim \mathcal{N}(0, (1-\alpha_j)P_{c,j})$ is independent of $X_{p,j}$ and

$$P_{c,j} \overset{\text{def}}{=} \frac{1}{2^{nR_c}} \sum_{j=1}^{2^{nR_c}} X_{c,j}^2, \quad P_{p,j} \overset{\text{def}}{=} \frac{1}{2^{nR_p}} \sum_{j=1}^{2^{nR_p}} X_{p,j}^2. \quad (54)$$

The parameter $\alpha_j \in [0,1]$ is chosen so that the resulting covariance $K_{X_{p,j},X_{c,j},Y_{s,j},Y_{p,j}}$ is the same as that induced by the code. Inequality labeled with (e) follows from Jensen's inequality, by choosing $\alpha \in [0,1]$ such that

$$\alpha P_c = \frac{1}{n}\sum_{j=1}^{n} \alpha_j P_{c,j}, \tag{55}$$

and from the fact that the power constraint $\|X_c^n\|^2 \leq nP_c$ implies that $\frac{1}{n}\sum_{j=1}^{n} P_{c,j} = P_c$.

Similarly, we can upper bound $h(\hat{Y}_p)$ as follows:

$$\begin{aligned}
h(\hat{Y}_p^n) &\overset{(a)}{=} \sum_{j=1}^{n} h(\hat{Y}_{p,j}|\hat{Y}_{p,1}^{j-1}), \\
&\overset{(b)}{\leq} \sum_{j=1}^{n} h(\hat{Y}_{p,j}), \\
&\overset{(c)}{\leq} \sum_{j=1}^{n} \frac{1}{2}\log(2\pi e \mathbb{E}[\hat{Y}_{p,j}^2]), \\
&\overset{(d)}{=} \sum_{j=1}^{n} \frac{1}{2}\log\left(\frac{2\pi e}{a^2}\left(P_{p,j} + 2\sqrt{\alpha_j P_{p,j} P_{c,j}} + P_{c,j} + 1\right)\right), \\
&\overset{(e)}{\leq} \frac{n}{2}\log\left(\frac{2\pi e}{a^2}\left((\sqrt{P_p} + \sqrt{\alpha P_c})^2 + (1-\alpha)P_c + 1\right)\right),
\end{aligned} \tag{56}$$

where (a) follows from the chain rule and (b) follows from the fact that conditioning does not increase entropy, (c) holds since the Gaussian distribution maximizes the differential entropy for a fixed variance, (d) follows from the same argument as in (51) and (e) comes from Jensen's inequality applied to the $\log(\cdot)$ and the $\sqrt{\cdot}$ functions.

Let $f(x) \overset{\text{def}}{=} x - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}x} + e^{\frac{2}{n}h(\widetilde{Z}^n)}\right)$ over $x \in \mathbb{R}$. Then, we can express the bound on our linear functional in (50) as

$$n(\mu R_p + R_c) \leq \mu h(\hat{Y}_p^n) + f(h(\widetilde{Y}^n|m_p)) - \frac{n}{2}\log(2\pi e) + \mu n \epsilon_{p,n} + n\epsilon_{s,n}. \tag{57}$$

Observe that as long as $\mu \leq 1$, $f(x)$ is increasing. Hence we can obtain a further upper bound by substituting inequalities (52) and (56) into (57):

$$
\begin{aligned}
n(\mu R_p + R_c) &\leq \mu \frac{n}{2} \log\left(\frac{2\pi e}{a^2}\left((\sqrt{P_p}+\sqrt{\alpha P_c})^2 + (1-\alpha)P_c + 1\right)\right) && (58)\\
&\quad + f\left(\frac{n}{2}\log\left(2\pi e\left((1-\alpha)P_c+1\right)\right)\right) - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, && (59)\\
&\overset{(a)}{=} \mu \frac{n}{2} \log\left(\frac{2\pi e}{a^2}\left((\sqrt{P_p}+\sqrt{\alpha P_c})^2 + (1-\alpha)P_c + 1\right)\right) && (60)\\
&\quad + \frac{n}{2}\log\left(2\pi e\left((1-\alpha)P_c+1\right)\right) - \mu\frac{n}{2}\log\left(2\pi e\left((1-\alpha)P_c + \frac{1}{a^2}\right)\right) && (61)\\
&\quad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}, && (62)
\end{aligned}
$$

where (a) follows from the fact that

$$
\begin{aligned}
f(x) &= x - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}x} + e^{\frac{2}{n}h(\widetilde{Z}^n)}\right), && (63)\\
&= x - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}x} + 2\pi e\left(\frac{1}{a^2}-1\right)\right), && (64)
\end{aligned}
$$

which holds since $\widetilde{Z}^n$ is zero mean Gaussian with covariance $\left(\frac{1}{a^2}-1\right)\mathbf{I}$.

Grouping together the μ-terms, dividing by n and letting $n \to \infty$, we get that

$$
\mu R_p + R_c \leq \frac{\mu}{2}\log\left(1 + \frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right) + \frac{1}{2}\log\left(1+(1-\alpha)P_c\right). \quad (65)
$$

Let α_μ denote the maximizing $\alpha \in [0,1]$ for a given $\mu \leq 1$ in the above expression. Then, we can write

$$
\mu R_p + R_c \leq \frac{\mu}{2}\log\left(1 + \frac{(\sqrt{P_p}+a\sqrt{\alpha_\mu P_c})^2}{1+a^2(1-\alpha_\mu)P_c}\right) + \frac{1}{2}\log\left(1+(1-\alpha_\mu)P_c\right). \quad (66)
$$

Hence we have established the converse of the theorem for $\mu \leq 1$.

A.2 The $\mu > 1$ case

A.2.1 Proof outline

Suppose that "genie A" gives the message m_p to the cognitive receiver. We will refer to this channel as the IC-DMS(A). The capacity region of the IC-DMS(A) must contain the capacity region of the original IC-DMS.

Proposition A.2 *The capacity region of the* $(1, a, 0, 1)$*-IC-DMS(A) is identical to the capacity region of* $(1, a, b, 1)$*-IC-DMS(A) for every* $b \in \mathbb{R}$ *and every* $a \in \mathbb{R}$.

Proof: Since m_p is known at the secondary receiver along with the primary encoding rule e_p^n, the secondary receiver of the $(1, a, 0, 1)$-IC-DMS(A) can form bX_p^n and add it to its received signal Y_s^n. The result is statistically identical to the the output at the secondary receiver of the $(1, a, b, 1)$-IC-DMS(A). Thus the capacity region is independent of b. □

This proposition allows us to set $b = 0$ without loss of generality in any IC-DMS(A).

Now suppose that "genie B" gives m_c to the primary transmitter of the $(1, a, 0, 1)$-IC-DMS(A). We will refer to this channel as the $(1, a, 0, 1)$-IC-DMS(A,B) and we note that its capacity region must contain the capacity region of the original $(1, a, b, 1)$-IC-DMS as well as that of the IC-DMS(A). Observe that this channel is equivalent to a broadcast channel with two antennas at the transmitter and one antenna at each of the receivers (2×1 MIMO BC channel) with per-antenna power constraints but *with additional knowledge* of m_p at the secondary receiver.

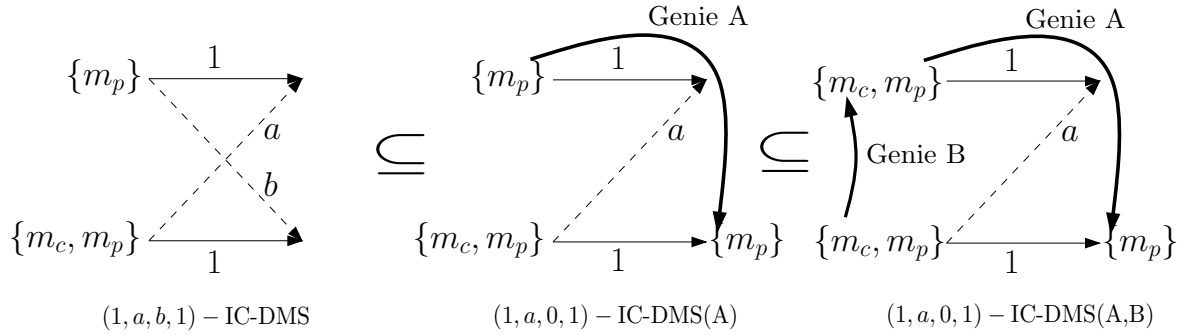

Figure 4: The $(1, a, b, 1)$-IC-DMS, the $(1, a, 0, 1)$-IC-DMS(A) and the $(1, a, 0, 1)$-IC-DMS(A,B) channels and the relationships between their capacity regions.

Thus, if we can show that the rates achieved by our proposed scheme for the $(1, a, b, 1)$-IC-DMS (given by (24) and (25)) are optimal for the $(1, a, 0, 1)$-IC-DMS(A,B), then we are done. To this end, we will first define a sequence of channels – each of which has a capacity region that includes the capacity region of the $(1, a, 0, 1)$-IC-DMS(A,B) – such that the rates (24) and (25) are optimal in the limit.

A.2.2 The aligned $(1, a, 0, 1)$-IC-DMS(A,B): The achievability

Consider the following modification of the $(1, a, 0, 1)$-IC-DMS(A,B): Add one antenna at each of the receivers so that the input-output relationship becomes

$$\mathbf{Y}_p = \begin{bmatrix} 1 & a \\ 1 & 0 \end{bmatrix} \mathbf{X} + \mathbf{Z}_p, \tag{67}$$

$$\mathbf{Y}_s = \begin{bmatrix} \epsilon & 1 \\ 0 & 1 \end{bmatrix} \mathbf{X} + \mathbf{Z}_s, \tag{68}$$

where $\epsilon > 0$ and $a \neq 0$. The vectors $\mathbf{Z}_p$ and $\mathbf{Z}_s$ are distributed according to $\mathcal{N}(0, \mathbf{\Sigma}_z)$ (their cross-correlation is irrelevant), where

$$\mathbf{\Sigma}_z = \begin{bmatrix} 1 & 0 \\ 0 & M \end{bmatrix}, \tag{69}$$

for some $M > 0$. As in the original $(1, a, 0, 1)$-IC-DMS(A,B), the message m_p is known at the secondary receiver. Clearly, the capacity region of this channel contains the capacity region of the $(1, a, 0, 1)$-IC-DMS(A,B). We shall refer to this genie-aided MIMO BC channel as the *aligned* $(1, a, 0, 1)$-IC-DMS(A,B) in what follows.

Let $\mathbf{H}_p$ and $\mathbf{H}_s$ denote the matrices pre-multiplying the transmit vector $\mathbf{X}$ in (67) and (68), respectively. Each coordinate of the vector $\mathbf{X} \in \mathbb{R}^2$ represents the symbol on each of the antennas and the constraint on $\mathbf{X}$ can in general take the form $\mathbb{E}[\mathbf{X}\mathbf{X}^T] \preceq \mathbf{Q}$ for some positive semi-definite covariance constraint $\mathbf{Q} \succeq 0$. Let the transmitted vector (at any time-sample) be of the form

$$\mathbf{X} = X_{p1}\mathbf{u}_{p1} + X_{p2}\mathbf{u}_{p2} + X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2}, \tag{70}$$

where $\mathbf{u}_{p1}, \mathbf{u}_{p2} \in \mathbb{R}^2$ and $\mathbf{u}_{c1}, \mathbf{u}_{c2} \in \mathbb{R}^2$ are the so-called signature vectors and symbols X_{p1}, X_{p2} and X_{c1}, X_{c2} are i.i.d. $\mathcal{N}(0, 1)$.

In order to emulate the per-user individual power constraints of the IC-DMS, we impose the per-antenna constraints $(\mathbb{E}[\mathbf{X}\mathbf{X}^T])_{11} \leq P_p$ and $(\mathbb{E}[\mathbf{X}\mathbf{X}^T])_{22} \leq P_c$ on the achievable strategies in MIMO BC channel. We let

$$\mathbf{\Sigma}_p \stackrel{\text{def}}{=} \mathbf{u}_{p1}\mathbf{u}_{p1}^T + \mathbf{u}_{p2}\mathbf{u}_{p2}^T, \tag{71}$$

$$\mathbf{\Sigma}_c \stackrel{\text{def}}{=} \mathbf{u}_{c1}\mathbf{u}_{c1}^T + \mathbf{u}_{c2}\mathbf{u}_{c2}^T, \tag{72}$$

so that, by the independence of X_{p1}, X_{p2}, X_{c1} and X_{c2}, the constraint can be expressed as $(\mathbf{\Sigma}_p + \mathbf{\Sigma}_c)_{11} \leq P_p$ and $(\mathbf{\Sigma}_p + \mathbf{\Sigma}_c)_{22} \leq P_c$.

Substituting the expression for $\mathbf{X}$ given in (70), the channel equations become

$$\mathbf{Y}_p = \mathbf{H}_p(X_{p1}\mathbf{u}_{p1} + X_{p2}\mathbf{u}_{p2}) + \mathbf{H}_p(X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2}) + \mathbf{Z}_p, \tag{73}$$

$$\mathbf{Y}_s = \mathbf{H}_s(X_{p1}\mathbf{u}_{p1} + X_{p2}\mathbf{u}_{p2}) + \mathbf{H}_s(X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2}) + \mathbf{Z}_s. \tag{74}$$

Consider the following encoding scheme: first choose X_{p1} and X_{p2} to be independent and distributed according to $\mathcal{N}(0, 1)$, and then perform Costa precoding to encode the information in (X_{c1}, X_{c2}) treating the interference $\mathbf{H}_s(X_{p1}\mathbf{u}_{p1} + X_{p2}\mathbf{u}_{p2})$ as side-information known at the transmitter[12]. The rates achievable with such a scheme are:

$$R_p = R_p(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \stackrel{\text{def}}{=} \frac{1}{2}\log\left|\mathbf{I} + (\mathbf{I} + \mathbf{\Sigma}_z^{-1}\mathbf{H}_p\mathbf{\Sigma}_c^*\mathbf{H}_p^T)^{-1}\mathbf{\Sigma}_z^{-1}\mathbf{H}_p\mathbf{\Sigma}_p^*\mathbf{H}_p^T\right|, \tag{75}$$

$$R_c = R_c(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \stackrel{\text{def}}{=} \frac{1}{2}\log\left|\mathbf{I} + \mathbf{\Sigma}_z^{-1}\mathbf{H}_s\mathbf{\Sigma}_c^*\mathbf{H}_s^T\right|, \tag{76}$$

[12]Costa's scheme is a block-coding scheme and, strictly speaking, encoding is performed on the vector (X_{c1}^n, X_{c2}^n) given X_{p1}^n and X_{p2}^n.

where $\mathbf{\Sigma}_p^*$ and $\mathbf{\Sigma}_c^*$ are the solutions of

$$\arg \max_{(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) \in \mathcal{S}(P_p, P_c)} \mu R_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) + R_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c), \tag{77}$$

where $\mu > 1$ and $\mathcal{S}(P_p, P_c) \stackrel{\text{def}}{=} \{\mathbf{\Sigma}_p \succeq 0, \mathbf{\Sigma}_c \succeq 0 : (\mathbf{\Sigma}_p + \mathbf{\Sigma}_c)_{11} \leq P_p,\ (\mathbf{\Sigma}_p + \mathbf{\Sigma}_c)_{22} \leq P_c\}$.

Since the per-antenna power constraints must be met with equality,[13] we can, without loss of generality, write

$$\mathbf{\Sigma}_p = \begin{bmatrix} \beta P_p & k_p \\ k_p & \alpha P_c \end{bmatrix}, \quad \text{where } k_p \in \left[-\sqrt{\alpha\beta P_p P_c}, \sqrt{\alpha\beta P_p P_c}\right], \tag{78}$$

$$\mathbf{\Sigma}_c = \begin{bmatrix} (1-\beta)P_p & k_c \\ k_c & (1-\alpha)P_c \end{bmatrix}, \quad \text{where } k_c \in \left[-\sqrt{\bar{\alpha}\bar{\beta} P_p P_c}, \sqrt{\bar{\alpha}\bar{\beta} P_p P_c}\right], \tag{79}$$

and $\beta \in [0,1]$, $\alpha \in [0,1]$ and $\bar{\alpha} \stackrel{\text{def}}{=} 1-\alpha$, $\bar{\beta} \stackrel{\text{def}}{=} 1-\beta$. With $\mathbf{\Sigma}_c$ expressed in this way, we obtain

$$\lim_{M\to\infty} \lim_{\epsilon \to 0} \mathbf{\Sigma}_z^{-1} \mathbf{H}_s \mathbf{\Sigma}_c \mathbf{H}_s^T = \begin{bmatrix} (1-\alpha)P_c & (1-\alpha)P_c \\ 0 & 0 \end{bmatrix}, \tag{80}$$

in (76). Similarly, by direct matrix calculations we get

$$\lim_{M\to\infty} \lim_{\epsilon \to 0} (\mathbf{I} + \mathbf{\Sigma}_z^{-1} \mathbf{H}_p \mathbf{\Sigma}_c \mathbf{H}_p^T)^{-1} = \begin{bmatrix} \frac{1}{(1-\beta)P_p + 2ak_c + a^2(1-\alpha)P_c + 1} & \frac{-(1-\beta)P_c - ak_c}{(1-\beta)P_p + 2ak_c + a^2(1-\alpha)P_c + 1} \\ 0 & 1 \end{bmatrix} \tag{81}$$

$$\lim_{M\to\infty} \lim_{\epsilon \to 0} \mathbf{\Sigma}_z^{-1} \mathbf{H}_p \mathbf{\Sigma}_p \mathbf{H}_p^T = \begin{bmatrix} \beta P_p + 2ak_p + a^2 \alpha P_c & \beta P_p + ak_p \\ 0 & 0 \end{bmatrix}. \tag{82}$$

Hence, on the one hand we have, by the continuity of $R_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c)$ in M and ϵ, that

$$\lim_{M\to\infty} \lim_{\epsilon \to 0} R_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2} \log(1 + (1-\alpha)P_c), \tag{83}$$

for any choice of $\beta \in [0,1]$. On the other hand, we have, by the continuity of $R_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c)$ in M and ϵ, that

$$\lim_{M\to\infty} \lim_{\epsilon \to 0} R_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2} \log\left(1 + \frac{\beta P_p + 2ak_p + a^2 \alpha P_c}{(1-\beta)P_p + 2ak_c + a^2(1-\alpha)P_c + 1}\right). \tag{84}$$

The limiting rate (84) is maximized by choosing $\beta = 1$ (and, therefore, $k_c = 0$) and $k_p = \sqrt{\alpha P_p P_c}$. Thus,

$$\mathbf{\Sigma}_p^* = \begin{bmatrix} P_p & \sqrt{\alpha P_p P_c} \\ \sqrt{\alpha P_p P_c} & \alpha P_c \end{bmatrix}, \tag{85}$$

$$\mathbf{\Sigma}_c^* = \begin{bmatrix} 0 & 0 \\ 0 & (1-\alpha)P_c \end{bmatrix}, \tag{86}$$

[13] If, instead, antenna 1 uses only $P_p - \eta$ power, we can add another antenna with power η whose signal the receivers can first decode and then subtract off thus boosting at least one of the rates. The same applies to antenna 2.

which is achieved by simply choosing

$$\mathbf{u}_{p1}^* = \begin{bmatrix} \sqrt{P_p} \\ \sqrt{\alpha P_c} \end{bmatrix}, \quad \mathbf{u}_{c1}^* = \begin{bmatrix} 0 \\ \sqrt{(1-\alpha)P_c} \end{bmatrix}, \quad \mathbf{u}_{p2}^* = \mathbf{0}, \quad \mathbf{u}_{c2}^* = \mathbf{0}. \tag{87}$$

Therefore, in the limit as $M \to \infty$ and $\epsilon \to 0$, this scheme achieves the rates given by (24) and (25) in the aligned $(1, a, 0, 1)$-IC-DMS(A,B).

A.2.3 The aligned $(1, a, 0, 1)$-IC-DMS(A,B): The converse

Since both $\mathbf{H}_p$ and $\mathbf{H}_s$ are invertible for every $\epsilon > 0$ and $a \neq 0$, we can equivalently represent this channel by the equations

$$\widetilde{\mathbf{Y}}_p = \mathbf{X} + \widetilde{\mathbf{Z}}_p, \tag{88}$$

$$\widetilde{\mathbf{Y}}_s = \mathbf{X} + \widetilde{\mathbf{Z}}_s. \tag{89}$$

The new noise vectors are given by $\widetilde{\mathbf{Z}}_p \sim \mathcal{N}(0, \mathbf{H}_p^{-1}\mathbf{\Sigma}_z\mathbf{H}_p^{-T})$ and $\widetilde{\mathbf{Z}}_s \sim \mathcal{N}(0, \mathbf{H}_s^{-1}\mathbf{\Sigma}_z\mathbf{H}_s^{-T})$. This channel is then exactly in the form of an Aligned MIMO BC channel (AMBC) (see [19], Section 2), but with m_p revealed to the secondary receiver.

Let $\widetilde{\mathbf{Y}}_p^n \in \mathbb{R}^{2\times n}$ and $\widetilde{\mathbf{Y}}_s^n \in \mathbb{R}^{2\times n}$ denote the channel outputs over a block of n channel uses. We can upper bound any achievable rate R_p as follows

$$nR_p = H(m_p), \tag{90}$$

$$= I(m_p; \widetilde{\mathbf{Y}}_p^n) + H(m_p|\widetilde{\mathbf{Y}}_p^n), \tag{91}$$

$$\overset{(a)}{\leq} I(m_p; \widetilde{\mathbf{Y}}_p^n) + n\widetilde{\epsilon}_{p,n}, \tag{92}$$

where (a) follows from Fano's inequality with $\widetilde{\epsilon}_{p,n} \to 0$ as $n \to \infty$. Noting that the secondary receiver observes the tuple $(\widetilde{\mathbf{Y}}_s^n, m_p)$, we can write

$$nR_c = H(m_c), \tag{93}$$

$$= H(m_c) + H(m_c|(\widetilde{\mathbf{Y}}_s^n, m_p)) - H(m_c|(\widetilde{\mathbf{Y}}_s^n, m_p)), \tag{94}$$

$$= I(m_c; (\widetilde{\mathbf{Y}}_s^n, m_p)) + H(m_c|(\widetilde{\mathbf{Y}}_s^n, m_p)), \tag{95}$$

$$\overset{(a)}{\leq} I(m_c; (\widetilde{\mathbf{Y}}_s^n, m_p)) + n\widetilde{\epsilon}_{s,n}, \tag{96}$$

$$\overset{(b)}{=} I(m_c; \widetilde{\mathbf{Y}}_s^n|m_p) + n\widetilde{\epsilon}_{s,n}, \tag{97}$$

where (a) follows from Fano's inequality with $\widetilde{\epsilon}_{s,n} \to 0$ as $n \to \infty$, and (b) follows since $I(m_p; m_c) = 0$.

Thus, we can upper-bound the linear functional of the achievable rates as

$$\mu R_p + R_c \leq \frac{\mu}{n} I(m_p; \widetilde{\mathbf{Y}}_p^n) + \frac{1}{n} I(m_c; \widetilde{\mathbf{Y}}_s^n|m_p) + \mu\widetilde{\epsilon}_{p,n} + \widetilde{\epsilon}_{s,n},$$

$$= \frac{\mu}{n} h(\widetilde{\mathbf{Y}}_p^n) - \frac{\mu}{n} h(\widetilde{\mathbf{Y}}_p^n|m_p) + \frac{1}{n} h(\widetilde{\mathbf{Y}}_s^n|m_p) - \frac{1}{n} h(\widetilde{\mathbf{Z}}_s^n) + \mu\widetilde{\epsilon}_{p,n} + \widetilde{\epsilon}_{s,n} \tag{98}$$

where $\mu > 1$.

Now, from Proposition 4.2 of [19], we know that, for every $\mu > 1$, there exists an *enhanced* Aligned Degraded BC channel (ADBC) which contains the capacity region of the AMBC given by (88) and (89), and for which the maximum of the linear functional $\mu R_p + R_c$, over all (R_p, R_c) in the region, is equal to the maximum of the same linear functional over the capacity region of the corresponding AMBC (i.e., the two regions meet at the point of tangency). Due to the degradedness, we can write the channel outputs of the enhanced ADBC as

$$\begin{aligned} \bar{\mathbf{Y}}_s^n &= \mathbf{X}^n + \bar{\mathbf{Z}}_s^n, && (99)\\ \bar{\mathbf{Y}}_p^n &= \bar{\mathbf{Y}}_s^n + \bar{\mathbf{Z}}_p^n, && (100)\\ & && (101) \end{aligned}$$

where the matrices $\bar{\mathbf{Z}}_s^n$ and $\bar{\mathbf{Z}}_p^n$ are constructed such that their columns, denoted by $\bar{\mathbf{Z}}_s$ and $\bar{\mathbf{Z}}_p$, are independent, zero-mean Gaussian with covariances satisfying $\mathbf{\Sigma}_{\bar{\mathbf{Z}}_s} \preceq \mathbf{\Sigma}_{\tilde{\mathbf{Z}}_s}$ and $\mathbf{\Sigma}_{\bar{\mathbf{Z}}_s} + \mathbf{\Sigma}_{\bar{\mathbf{Z}}_p} \preceq \mathbf{\Sigma}_{\tilde{\mathbf{Z}}_p}$ (see proof of Proposition 4.2 of [19] for how to construct them). Hence, for this enhanced ADBC, we can write (98) as

$$\begin{aligned} \mu R_p + R_c &\leq \frac{\mu}{n} h(\bar{\mathbf{Y}}_p^n) + \frac{1}{n} h(\bar{\mathbf{Y}}_s^n | m_p) - \frac{\mu}{n} h(\bar{\mathbf{Y}}_p^n | m_p) - \frac{1}{n} h(\bar{\mathbf{Z}}_s^n) + \mu \bar{\epsilon}_n \\ &= \frac{\mu}{n} h(\bar{\mathbf{Y}}_s^n + \bar{\mathbf{Z}}_p^n) + \frac{1}{n} h(\bar{\mathbf{Y}}_s^n | m_p) - \frac{\mu}{n} h(\bar{\mathbf{Y}}_s^n + \bar{\mathbf{Z}}_p^n | m_p) - \frac{1}{n} h(\bar{\mathbf{Z}}_s^n) + \mu \bar{\epsilon}_n, \\ &\leq \frac{\mu}{n} h(\bar{\mathbf{Y}}_p^n) + \frac{1}{n} h(\bar{\mathbf{Y}}_s^n | m_p) - \mu \log \left(e^{\frac{2}{2n} h(\bar{\mathbf{Y}}_s^n | m_p)} + e^{\frac{2}{2n} h(\bar{\mathbf{Z}}_p^n)} \right) \\ &\qquad - \frac{1}{n} h(\bar{\mathbf{Z}}_s^n) + \mu \bar{\epsilon}_n, \quad (102) \end{aligned}$$

where we have used the conditional version of the vector Entropy Power Inequality (see Proposition D.1) in the last step.

The key property of the this enhanced ADBC is that the upper bound (102) is maximized by choosing the input $\mathbf{X}$ to be Gaussian, i.e., the vector EPI is tight (see proof of Theorem 3.1 of [19]). Hence, an optimal achievable scheme for this ADBC is the Costa precoding strategy[14] that is described in Section A.2.2: The largest jointly achievable rates are given by

$$\begin{aligned} R_p &= R_p(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*), && (103)\\ R_c &= R_c(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) && (104) \end{aligned}$$

where $R_p(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*)$ and $R_c(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*)$ are as given by (75) and (76), respectively.

Since this scheme is also achievable for the AMBC, the capacity region of the ADBC and AMBC are identical (see Theorem 4.1 of [19]). Moreover, it is obvious that this scheme is also achievable for the AMBC with *additional knowledge* of m_p at the secondary

[14]Note that for the ADBC a simple superposition scheme is also optimal.

receiver: The knowledge of m_p is simply ignored by the receiver. Hence, this scheme is optimal for the aligned $(1, a, 0, 1)$-IC-DMS(A,B) (as defined by (67) and (68)) with $\mu > 1$ as well.

Since the Pareto-optimal (for $\mu > 1$) rates for the limiting (as $M \to \infty$ and $\epsilon \to 0$) aligned $(1, a, 0, 1)$-IC-DMS(A,B) exactly match the rates (24) and (25) achievable in the original $(1, a, b, 1)$-IC-DMS channel, *and* since the capacity region of the $(1, a, b, 1)$-IC-DMS is contained in the capacity region of the aligned $(1, a, 0, 1)$-IC-DMS(A,B) for any $M, \epsilon > 0$, we have completed the proof of the converse part of Theorem 4.1 for $\mu > 1$.

B Proof of Corollary 4.1

The proof of this Corollary follows from Theorem 4.1 and Lemma D.2. In particular, we observe that the converse to Theorem 4.1 for $\mu \geq 1$ (see Section A.2) holds for any $a > 0$ and $b \in \mathbb{R}$. However, from Lemma D.2 we see that the choice $\alpha = 1$ in (24) and (25) is optimal for any $a \geq 1$, as long as $\mu \geq 1$. Hence the corollary is proved. $\square$.

Remark: This result implies that, for any $a \geq 1$, $b \in \mathbb{R}$ and $\mu \geq 1$, the linear functional $\mu R_p + R_c$ is maximized at $(R_p, R_c) = (C_{\text{sum}}(a), 0)$. Hence, for $a \geq 1$, the entire capacity region is parametrized by $\mu \leq 1$, for any $b \in \mathbb{R}$.

C Proof of the converse part of Theorem 4.3

Let "genie B" disclose m_c to the primary transmitter, thus getting a 2×1 MIMO BC channel with per-antenna power constraints. The input-output relationship for this channel can be written as

$$Y_p = \mathbf{h}_p^T \mathbf{X} + Z_p, \tag{105}$$
$$Y_s = \mathbf{h}_s^T \mathbf{X} + Z_s, \tag{106}$$

where $\mathbf{h}_p = [1 \;\; a]^T$ and $\mathbf{h}_s = [b \;\; 1]^T$. We choose $\mu \leq 1$ in the linear functional $\mu R_p + R_c$ and recall that the optimal transmission vector $\mathbf{X}$ is Gaussian and given by (70) and the optimal encoding strategy is to generate X_p by Costa precoding for $\mathbf{h}_p^T(X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2})$ (see [19]). Consequently, in place of (75) and (76), we get, respectively,

$$R_p = \hat{R}_p(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \mathbf{h}_p^T \mathbf{\Sigma}_p^* \mathbf{h}_p\right), \tag{107}$$
$$R_c = \hat{R}_c(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \frac{\mathbf{h}_s^T \mathbf{\Sigma}_c^* \mathbf{h}_s}{1 + \mathbf{h}_s^T \mathbf{\Sigma}_p^* \mathbf{h}_s}\right), \tag{108}$$

where $\mathbf{\Sigma}_c^*$ and $\mathbf{\Sigma}_c^*$ are the solutions of (77) but with $\mu \leq 1$. Substituting the covariance matrices (78) and (79) into (107) and (108), we get

$$\hat{R}_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \hat{R}_p(\beta, \alpha, k_p, a, b) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \beta P_p + 2ak_p + \alpha a^2 P_c\right), \tag{109}$$

$$\hat{R}_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \hat{R}_c(\beta, \alpha, k_p, a, b) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \frac{b^2(1-\beta)P_p + 2k_c b + (1-\alpha)P_c}{1 + b^2\beta P_p + 2k_p b + \alpha P_c}\right) \tag{110}$$

The expression in (110) is maximized by choosing $k_c = \sqrt{(1-\beta)(1-\alpha)P_p P_c}$, i.e., making $\mathbf{\Sigma}_c$ unit rank. If $b = 0$ it is clear that $\beta = 1$ and $k_p = \sqrt{\alpha P_p P_c}$ maximizes the linear functional $\mu\hat{R}_p(\beta, \alpha, k_p, a, b) + \hat{R}_c(\beta, \alpha, k_p, a, b)$. In general, we would like to find the set of all values of b for which $\beta = 1$ and $k_p = \sqrt{\alpha P_p P_c}$ are optimal. For such values of b, we then have

$$\hat{R}_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2}\log\left(1 + \left(\sqrt{P_p} + a\sqrt{\alpha P_c}\right)^2\right), \tag{111}$$

$$\hat{R}_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2}\log\left(1 + \frac{(1-\alpha)P_c}{1 + \left(b\sqrt{P_p} + \sqrt{\alpha P_c}\right)^2}\right), \tag{112}$$

which exactly match the achievable rates given in Lemma 4.2. To this end, let $B(\mu, a)$ denote the set of all $b > 0$ such that the function

$$\max_{0\leq\alpha\leq 1} \mu\hat{R}_p\left(\beta, \alpha, k_p, a, b\right) + \hat{R}_c(\beta, \alpha, k_p, a, b) \tag{113}$$

is maximized, over all $\beta \in [0, 1]$ and $k_p \in [-\sqrt{\beta\alpha P_p P_c}, \sqrt{\beta\alpha P_p P_c}]$, by choosing $\beta = 1$ and $k_p = \sqrt{\alpha P_p P_c}$. We let $b_{\max}(\mu, a) \stackrel{\text{def}}{=} \max_{b\in B(\mu,a)}$ to obtain the statement of the theorem. Appealing to the remark in the proof of Corollary 4.1 (see Appendix B), we observe that the boundary of the capacity region in this very-high-interference-gain regime is completely parametrized by $\mu \leq 1$. Hence, we have proved the theorem.

D Supporting results

Proposition D.1 *The rate region specified in Theorem 4.1 is a convex set.*

Proof: A point $\mathbf{R} = (R_p, R_c)$ is in the rate region specified in Theorem 4.1 if and only if there exists $\alpha \in [0, 1]$ such that

$$0 \leq R_c \leq \frac{1}{2}\log(1 + (1-\alpha)P_c), \tag{114}$$

$$0 \leq R_p \leq \frac{1}{2}\log\left(1 + a^2 P_c + P_p + 2a\sqrt{\alpha P_p P_c}\right) + \frac{1}{2}\log\left(\frac{1}{1 + a^2(1-\alpha)P_c}\right). \tag{115}$$

Suppose that there exist two points $\mathbf{R}^{(1)} = (R_p^{(1)}, R_c^{(1)})$ and $\mathbf{R}^{(2)} = (R_p^{(2)}, R_c^{(2)})$ that are in the region. Let $\alpha^{(1)} \in [0,1]$ and $\alpha^{(2)} \in [0,1]$ be their corresponding parameters in (114) and (115). Then for any $\lambda \in [0,1]$, we have that

$$\begin{aligned}
\lambda R_c^{(1)} + (1-\lambda) R_c^{(2)} &\leq \frac{\lambda}{2}\log(1+(1-\alpha^{(1)})P_c) + \frac{1-\lambda}{2}\log(1+(1-\alpha^{(2)})P_c), && (116)\\
&\leq \frac{1}{2}\log(1+(1-\alpha^*)P_c) && (117)
\end{aligned}$$

where $\alpha^* \stackrel{\text{def}}{=} \lambda\alpha^{(1)} + (1-\lambda)\alpha^{(2)}$ and the last inequality follows from Jensen's inequality. Similarly,

$$\begin{aligned}
\lambda R_p^{(1)} + (1-\lambda) R_p^{(2)} &\leq \left[\frac{\lambda}{2}\log\left(1 + a^2 P_c + P_p + 2a\sqrt{\alpha^{(1)} P_p P_c}\right)\right. && (118)\\
&\quad \left. + \frac{1-\lambda}{2}\log\left(1 + a^2 P_c + P_p + 2a\sqrt{\alpha^{(2)} P_p P_c}\right)\right] && (119)\\
&\quad + \left[\frac{\lambda}{2}\log\left(\frac{1}{1+a^2(1-\alpha^{(1)})P_c}\right)\right. && (120)\\
&\quad \left. + \frac{1-\lambda}{2}\log\left(\frac{1}{1+a^2(1-\alpha^{(2)})P_c}\right)\right], && (121)\\
&\stackrel{(a)}{\leq} \frac{1}{2}\log\left(1 + a^2 P_c + P_p + 2a\sqrt{P_p P_c}\left(\lambda\sqrt{\alpha^{(1)}} + (1-\lambda)\sqrt{\alpha^{(2)}}\right)\right) && (122)\\
&\quad + \frac{1}{2}\log\left(\frac{1}{1+a^2(1-\lambda\alpha^{(1)} - (1-\lambda)\alpha^{(2)})P_c}\right), && (123)\\
&\stackrel{(b)}{\leq} \frac{1}{2}\log\left(1 + a^2 P_c + P_p + 2a\sqrt{P_p P_c \alpha^*}\right) + \frac{1}{2}\log\left(\frac{1}{1+a^2(1-\alpha^*)P_c}\right) && (124)
\end{aligned}$$

(a) follows from Jensen's inequality applied to the concave function $\log(k_1 + k_2 x)$ (for constant $k_1, k_2 > 0$) and the concave function $\log\left(\frac{1}{1+(1-x)k}\right)$ (for constant $k > 0$). Inequality (b) follows from Jensen's inequality applied to the square-root function. Hence $\lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$ is in the region as well, hence the region is a convex set. □

Proposition D.2 (Conditional EPI) *Suppose $Y^n \in \mathbb{R}^n$ and $Z^n \in \mathbb{R}^n$ are independent random vectors and $m \in \{1, 2, \ldots, M\}$ (for some M) is independent of Z^n. Then we have that*

$$h(Y^n + Z^n | m) \geq \frac{n}{2}\log\left(e^{\frac{2}{n}h(Y^n|m)} + e^{\frac{2}{n}h(Z^n)}\right). \tag{125}$$

Proof:

$$\begin{aligned} h(Y^n + Z^n|m) &= \sum_{i=1}^{M} h(Y^n + Z^n|m=i)\mathbb{P}(m=i), && (126)\\ &\overset{(a)}{\geq} \sum_{i=1}^{M} \frac{n}{2}\log\left(e^{\frac{2}{n}h(Y^n|m=i)} + e^{\frac{2}{n}h(Z^n)}\right)\mathbb{P}(m=i), && (127)\\ &\overset{(b)}{\geq} \frac{n}{2}\log\left(e^{\frac{2}{n}h(Y^n|m)} + e^{\frac{2}{n}h(Z^n)}\right), && (128) \end{aligned}$$

where (a) follows from the classical Entropy Power Inequality (EPI) (see e.g. [5]), and (b) follows from Jensen's inequality applied to the convex function $\log(e^{2x/n} + k)$ (for constant k and n). □.

Lemma D.1 *Given two zero-mean random variables X and Y with a fixed covariance matrix K_{XY} we have that*

$$h(Y|X) \leq \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}[Y^2] - \frac{\mathbb{E}[YX]^2}{\mathbb{E}[X^2]}\right)\right), \quad (129)$$

with equality when X and Y are jointly Gaussian.

Proof: Let $\beta = \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}$. Then the MMSE estimator of Y given X is given by $\hat{Y} = \beta X$.

$$\begin{aligned} h(Y|X) &\overset{(a)}{=} h(Y - \beta X|X), && (130)\\ &\overset{(b)}{\leq} h(Y - \beta X), && (131)\\ &\overset{(c)}{\leq} \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}[(Y-\beta X)^2]\right)\right), && (132)\\ &= \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}[Y^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[X^2]}\right)\right), && (133) \end{aligned}$$

where (a) follows from the fact that shifts do not change the differential entropy, (b) follows since conditioning does not increase entropy, and (c) follows since the Gaussian distribution maximizes the entropy for a given variance. By the orthogonality principle, (b) is tight when X and Y are jointly Gaussian and in that case (c) is tight as well. □

Lemma D.2

$$\begin{aligned} \max_{0\leq\alpha\leq 1} \frac{\mu}{2}\log\left(1 + \frac{(\sqrt{P_p} + a\sqrt{\alpha P_c})^2}{1 + a^2(1-\alpha)P_c}\right) + \frac{1}{2}\log\left(1 + (1-\alpha)P_c\right) \quad (134)\\ = \frac{\mu}{2}\log\left(1 + \left(\sqrt{P_p} + a\sqrt{P_c}\right)^2\right), \end{aligned}$$

for $a \geq 1$ and $\mu \geq 1$.

Proof: On the one hand we have that

$$\max_{0\leq\alpha\leq 1} \frac{\mu}{2}\log\left(1+\frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right)+\frac{1}{2}\log\left(1+(1-\alpha)P_c\right), \tag{135}$$

$$= \max_{0\leq\alpha\leq 1} \frac{1}{2}\log\left(\frac{\left(1+a^2(1-\alpha)P_c+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2\right)^{\mu}\left(1+(1-\alpha)P_c\right)}{(1+a^2(1-\alpha)P_c)^{\mu}}\right), \tag{136}$$

$$\leq \max_{0\leq\alpha\leq 1} \frac{1}{2}\log\left(\frac{\left(1+a^2(1-\alpha)P_c+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2\right)^{\mu}}{(1+a^2(1-\alpha)P_c)^{\mu-1}}\right), \tag{137}$$

$$= \max_{0\leq\alpha\leq 1} \frac{1}{2}\log\left(\frac{\left(1+a^2P_c+P_p+2a\sqrt{\alpha P_pP_c}\right)^{\mu}}{(1+a^2(1-\alpha)P_c)^{\mu-1}}\right), \tag{138}$$

$$= \frac{\mu}{2}\log\left(1+\left(\sqrt{P_p}+a\sqrt{P_c}\right)^2\right). \tag{139}$$

On the other hand, the maximization problem in (134) can be lower bounded with $\frac{\mu}{2}\log\left(1+\left(\sqrt{P_p}+a\sqrt{P_c}\right)^2\right)$, by choosing $\alpha=1$. Hence the lemma is proved. □

References

[1] A.B. Carleial, "Interference channels," *IEEE Transactions on Information Theory,* vol. 24, no. 1, pp. 60-70, Jan. 1978

[2] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, pp. 16391667, June 2002.

[3] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[4] H.M. Costa, "On the Gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 607-615, Sept. 1985.

[5] A. Dembo, T.M. Cover, and J.A. Thomas, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1501-1518, Nov. 1991.

[6] N. Devroye, P. Mitran, and V. Tarokh, "Achievable Rates in Cognitive Channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813-1827, May 2006.

[7] Federal Communications Commission Spectrum Policy Task Force, "Report of the Spectrum Efficiency Working Group", FCC, Tech Report, Nov. 2002.

[8] Federal Communications Commision, Cognitive Radio Technologies Proceeding (CRTP), ET Docket No. 03-108, `http://www.fcc.gov/oet/cognitiveradio/`.

[9] IEEE Standards Association, "IEEE 802.16e Mobile WirelessMAN (R) Standard is Official" http://standards.ieee.org/announcements/pr_p80216.html, Dec. 2005.

[10] International Telecommunications Union draft, "Characteristics of the IEEE 802.16 systems in the $2500 - 2690$ MHz", Dec. 2004, http://wirelessman.org/liaison/docs/L80216-04_42r2.pdf

[11] A. Lapidoth, "Nearest-neighbor decoding for additive non-Gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1520-1529, Sept. 1996.

[12] I. Maric, R. D. Yates and G. Kramer, "The Capacity Region of the Strong Interference Channel With Common Information," Asilomar Conference On Signals, Systems and Computers, Pacific Grove, CA, Nov. 2005.

[13] P. Mitran, H. Ochiai, V. Tarokh, "Space-Time Diversity Enhancements Using Collaborative Communications," in *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2041-2057, June 2005.

[14] J. M. Peha and S. Panichpapiboon, "Real-Time Secondary Markets for Spectrum," Telecommunications Policy, vol. 28, pp. 603-618, Aug.-Sep. 2004

[15] R. T. Rockafellar *Convex Analysis*, Princeton University Press, 1971.

[16] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 786-788, Nov. 1981.

[17] A. Sutivong, M. Chiang, T. M. Cover, Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Transactions on Information Theory* vol. 51, no. 4, pp. 1486-1495, April 2005

[18] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", Cambridge University Press, 2005.

[19] H. Weingarten, Y. Steinberg, S. Shamai, "The Capacity Region of the Gaussian MIMO Broadcast Channel", submitted to *IEEE Transactions on Information Theory*, July 2004.