Cognitive Radio: An Information-Theoretic Perspective

Aleksandar Jovičić and Pramod Viswanath, *Member, IEEE*

***Abstract*—In this paper, we consider a communication scenario in which the primary and the cognitive radios wish to communicate to different receivers, subject to mutual interference. In the model that we use, the cognitive radio has noncausal knowledge of the primary radio's codeword. We characterize the largest rate at which the cognitive radio can reliably communicate under the constraint that 1) *no rate degradation* is created for the primary user, and 2) the primary receiver uses a *single-user decoder* just as it would in the absence of the cognitive radio. The result holds in a "low-interference" regime in which the cognitive radio is closer to its receiver than to the primary receiver. In this regime, our results are subsumed by the results derived in a concurrent and independent work (Wu *et al.*, 2007). We also demonstrate that, in a "high-interference" regime, multiuser decoding at the primary receiver is optimal from the standpoint of maximal jointly achievable rates for the primary and cognitive users.**

***Index Terms*—Cognitive radio, Costa precoding, dirty-paper coding, interference channel, spectral reuse, wireless networks.**

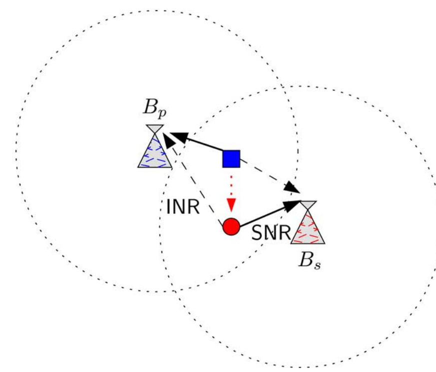

Fig. 1. A possible arrangement of the primary and secondary receivers, base stations B_p and B_s, respectively. The cognitive secondary user is represented by the circle and the primary user is represented by the square. The side-information path is depicted by the dotted line.

I. Introduction

OBSERVING a severe under-utilization of the licensed spectrum, the Federal Communications Commission (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 [16] 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. 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.

Manuscript received May 10, 2006; revised June 13, 2007. Current version published August 19, 2009. This work was supported in part by the National Science Foundation under Grant CCR-0312413 and by a grant from Motorola Inc. as part of the Motorola Center for Communication. The material in this paper was presented in part at the IEEE Symposium on Information Theory, Seattle, WA, July 2006.

A. Jovičić is with the Qualcomm Flarion Technologies, Bridgewater, NJ 08807 USA (e-mail: ajovicic@qualcomm.com).

P. Viswanath is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: pramodv@uiuc.edu).

Communicated by G. Kramer, Associate Editor for Special Issue: Relaying and Cooperation.

Color version of Figure 1 in this paper is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2025539

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 noncausal 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
> 1) it creates no rate degradation for the primary user in its vicinity, and
> 2) 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 SNR of this interfering signal is denoted by INR (interference-to-noise ratio).

[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.

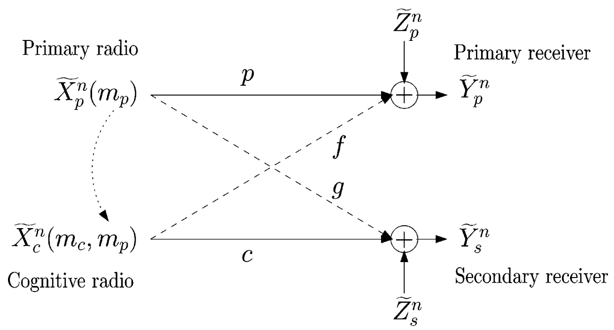

Fig. 2. The (Gaussian) cognitive radio channel after n channel uses. The dashed lines represent interfering receptions. The dotted line represents the side-information path. The power constraints are $\tilde{P}_p$ and $\tilde{P}_c$ and noise variances are N_p and N_s.

If the cognitive user is close to B_p, INR could potentially be large.

We identify 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 this result, we allow the primary and secondary systems to *cooperate* 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. Under such a joint encoder/decoder design, the cognitive radio channel can be thought of as the classical interference channel [1], [18], [4] but with degraded message sets[2]: Achievable schemes for this channel have been first studied in [6]. The capacity region of this channel, in the low-interference-gain regime, has recently been found by an independent and concurrent work [23], [24]. In this paper, we provide an alternative proof for a portion of the capacity region of the same channel. The capacity region under the assumption that both receivers experience high interference has been reported in [13]. A related problem of communicating a single private message along with a common message to each of the receivers has been studied in [14].

In contrast to our results for the low-interference-gain regime, we exhibit a regime in which joint code design *is* beneficial with respect to the largest set of simultaneously achievable rates of the two radios. We find that, when $\mathsf{INR} \gg \mathsf{SNR}$, *multiuser* decoding by the primary receiver is required in order to achieve all the rates in the capacity region of the interference channel with degraded message sets.

The rest of this paper is organized as follows. We first introduce the Gaussian *cognitive radio channel* in Section II. The capacity of the cognitive radio channel in the low-interference-gain regime, where $\mathsf{INR} \leq \mathsf{SNR}$, is presented in Section III. The proof is given in Section III, where we demonstrate a part of the capacity region of the underlying interference channel with degraded message sets which inherently allows for joint code design. This result, along with that of the parallel work by [24], yields the capacity of the Gaussian cognitive radio channel. 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 IV-B. Finally, we study the system-level implications of the optimal cognitive communication scheme in Section V.

II. The Channel Model and Problem Statement

A. The Cognitive Radio Channel

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

The additive noise at the primary and secondary receivers, $\tilde{Z}_p^n := (\tilde{Z}_{p,1}, \tilde{Z}_{p,2}, \ldots, \tilde{Z}_{p,n})$ and $\tilde{Z}_s^n := (\tilde{Z}_{s,1}, \tilde{Z}_{s,2}, \ldots, \tilde{Z}_{s,n})$, is assumed to be independent identically distributed (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 $\tilde{Z}_p^n$ and $\tilde{Z}_s^n$ 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 $\tilde{P}_p$ and $\tilde{P}_c$, respectively

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

The received signal-to-noise ratios (SNRs) of the desired signals at the primary and secondary base station are $p^2\tilde{P}_p/N_p$ and $c^2\tilde{P}_c/N_s$, respectively. The received SNRs of the interfering signals at the primary and secondary base station (INRs) are $f^2\tilde{P}_c/N_p$ and $g^2\tilde{P}_p/N_s$, 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 V-C). The channel can be described by the pair of per-time-sample equations

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

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

B. Transformation to Standard Form

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

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

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

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

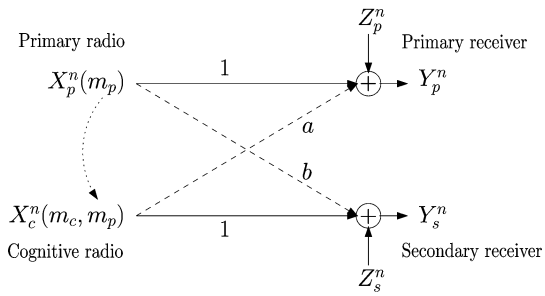

Fig. 3. The cognitive radio channel in standard form. The channel gains (p, f, g, c) in the original channel are mapped to $(1, a, b, 1)$, powers $(\tilde{P}_p, \tilde{P}_c)$ are mapped to (P_p, P_c), and noise variances (N_p, N_s) are mapped to $(1, 1)$.

where

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

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

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

We will consider this standard form of the cognitive radio channel without loss of generality and we will refer to it as the $(1, a, b, 1)$ *cognitive radio channel*.

C. Coding on the Cognitive Radio 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 *single-user decoder*, $D_p^n : \mathcal{Y}_p^n \mapsto \{1, 2, \ldots, 2^{nR_p}\}$, to decode m_p from Y_p^n. We define a single-user decoder to be any decoder which performs well on the point-to-point additive white Gaussian noise (AWGN) channel. For concreteness, we set the primary decoder to be the nearest-neighbor (or minimum-distance) decoder. Hence, we have the following definition.

Definition 2.1 (Achievability: Primary User): A rate R_p is achievable for the primary user if there exists a sequence of encoders $E_p^n : \{1, 2, \ldots, 2^{nR_p}\} \mapsto \mathcal{X}_p^n$ such that the average probability of error[5] vanishes as $n \to \infty$, i.e.,

$$P_{e,p}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{nR_p}} \sum_{i=1}^{2^{nR_p}} \mathbb{P}\left(D_p^n\left(Y_p^n\right) \neq i | m_p = i\right) \to 0 \tag{9}$$

when the primary receiver uses a single-user decoder D_p^n.

[4]These transformations were used in [1], [3], and [18], in the context of the classical interference channel.

[5]The tuple (m_p, m_c) is assumed to be drawn from a uniform product measure on $\{1, 2, \ldots, 2^{nR_p}\} \times \{1, 2, \ldots, 2^{nR_c}\}$.

The salient feature of the cognitive radio is that it has knowledge of the primary encoder as well as the message m_p. Let $\mathcal{E}_p^n$ denote the set of all primary encoders. Hence, we have the following definition.

Definition 2.2 (Cognitive Radio Code): A $(2^{nR_c}, n)$ cognitive radio code is a choice of an encoding rule (whose output we denote by X_c^n)

$$E_c^n : \mathcal{E}_p^n \times \{1, 2, \ldots, 2^{nR_p}\} \times \{1, 2, \ldots, 2^{nR_c}\} \to \mathcal{X}_c^n \tag{10}$$

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

$$D_c^n : \mathcal{Y}_s^n \to \{1, 2, \ldots, 2^{nR_c}\}. \tag{11}$$

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

Definition 2.3 (Achievability: Cognitive User): A rate R_c is said to be achievable for the cognitive user on a $(1, a, b, 1)$ cognitive radio channel if there exists a sequence of $(2^{nR_c}, n)$ cognitive radio codes such that the following two constraints are satisfied:

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

$$P_{e,c}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}\left(D_c^n\left(Y_s^n\right) \neq j | m_p = i, m_c = j\right) \to 0; \tag{12}$$

2) a rate of $R_p^* \stackrel{\text{def}}{=} \frac{1}{2}\log(1 + P_p)$ is achievable for the primary user, in the sense of Definition 2.1.

We note that the above definition is well defined, since the rate $R_c = 0$ is achievable for the cognitive user if the cognitive radio simply shuts off.

Definition 2.4 (Capacity): The *capacity* of the cognitive radio channel is defined to be the largest achievable rate R_c for the cognitive user.

III. The Capacity of the Cognitive Radio Channel

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 radio channel, or, equivalently, to $a \leq 1$ in the corresponding standard-form $(1, a, b, 1)$ cognitive radio channel. The capacity of the cognitive radio channel in this regime is given by the following expression.

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

$$R_c^* = \frac{1}{2}\log(1 + (1 - \alpha^*)P_c) \tag{13}$$

as long as $a \leq 1$. The constant $\alpha^* \in [0,1]$ is given by

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

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

A. Proof of 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 encoders and decoders.

- 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:

$$X_c^n = \hat{X}_c^n + \sqrt{\frac{\alpha P_c}{P_p}} X_p^n \tag{15}$$

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 noncausally 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_p^n: Single-user decoder that is optimal for the point-to-point AWGN channel, such as the nearest-neighbor decoder.
- 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

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

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

$$\frac{1}{2}\log(1 + (1-\alpha)P_c). \tag{17}$$

However, in order to ensure that a given rate is *achievable* for the cognitive user in the sense of Definition 2.3, we must have that

$$\frac{1}{2}\log\left(1 + \frac{(\sqrt{P_p} + a\sqrt{\alpha P_c})^2}{1 + a^2(1-\alpha)P_c}\right) = \frac{1}{2}\log(1+P_p) =: R_p^*. \tag{18}$$

Observe that, if $a = 0$, any choice of $\alpha \in [0,1]$ will satisfy (18): in this case, we should set $\alpha^* = 0$ to maximize the rate achievable for the cognitive user. For $a > 0$, by the intermediate value theorem, this quadratic equation in α always has a unique root in $[0,1]$, and it is given by (14).

Finally, since the code-ensemble-averaged (and message-averaged) probabilities of error vanish, there must exist a particular sequence of cognitive radio 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 2.3.

B. Proof of the Converse Part

1) Proof Outline: We will first relax the constraints of our problem and allow for *joint* primary and cognitive radio code design thus forming an interference channel with degraded message sets,[6] which we will abbreviate as IC-DMS for convenience. Since this relaxation enlarges the space of allowable encoder/decoder pairs, the largest set of achievable rate pairs (R_p, R_c) in the IC-DMS must include the rate point corresponding to the capacity of the cognitive radio channel, i.e., $(\frac{1}{2}\log(1+P_p), R_c^*)$, whatever R_c^* might be.

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 are able to do this only for a part of the capacity region[7] and we refer the reader to a concurrent and independent work [24] which provides a proof for the entire capacity region. Next we 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 largest value of R_c such that the point $(R_p, R_c) = (\frac{1}{2}\log(1+P_p), R_c)$ is on the boundary of the capacity region of the IC-DMS is given by $R_c^* = \frac{1}{2}\log(1 + (1-\alpha^*)P_c)$ with α^* as in (14). We then conclude that $R_c = R_c^*$ is the capacity of the corresponding cognitive radio channel.

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

Definition 3.1 (IC-DMS Code): A $(2^{nR_p}, 2^{nR_c}, n)$ code for the $(1, a, b, 1)$-IC-DMS is a joint selection of the encoding rules

[6]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.

[7]The original version of this paper had an error in the proof of the converse for the complementary portion of the capacity region.

and the decoding rules: the encoding rules are two maps (whose outputs we denote by X_p^n and X_c^n, respectively), denoted as

$$e_p^n : \{1, 2, \dots, 2^{nR_p}\} \to \mathcal{X}_p^n \quad (19)$$
$$e_c^n : \{1, 2, \dots, 2^{nR_p}\} \times \{1, 2, \dots, 2^{nR_c}\} \to \mathcal{X}_c^n \quad (20)$$

such that $\|X_p^n\|^2 \leq nP_p$ and $\|X_c^n\|^2 \leq nP_c$. The decoding rules are two maps denoted as

$$d_p^n : \mathcal{Y}_p^n \to \{1, 2, \dots, 2^{nR_p}\} \quad (21)$$
$$d_c^n : \mathcal{Y}_s^n \to \{1, 2, \dots, 2^{nR_c}\}. \quad (22)$$

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 $d_c^n(Y_s^n) \neq j$.

Observe that, unlike in the original cognitive radio channel, here we allow for an arbitrary choice of the primary decoder and, in particular, for the possibility of *multiuser* decoding, an example of which is successive interference decoding and cancellation.

Definition 3.2 (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.,

$$\tilde{P}_{e,p}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}(d_p^n(Y_p^n) \neq i | m_p = i, m_c = j) \to 0 \quad (23)$$
$$P_{e,c}^{(n)} \stackrel{\text{def}}{=} \frac{1}{2^{n(R_c+R_p)}} \sum_{i=1,j=1}^{n} \mathbb{P}(d_c^n(Y_s^n) \neq j | m_p = i, m_c = j) \to 0. \quad (24)$$

Definition 3.3 (Capacity Region): The capacity region of the IC-DMS is the closure of the set of achievable rate vectors (R_p, R_c).

3) The Capacity Region of the IC-DMS Under a Low-Interference Gain: The following theorem characterizes a part of the capacity region of the $(1, a, b, 1)$-IC-DMS with $a \leq 1$ and arbitrary $b \in \mathbb{R}$. As mentioned earlier, our result here is subsumed by the result independently and concurrently obtained in [24, Th. 3.5].

Theorem 3.2: The segment of the boundary, corresponding to $\frac{dR_c}{dR_p} \geq -1$, of the capacity region of the $(1, a, b, 1)$-IC-DMS with $a \leq 1$ and $b \in \mathbb{R}$ is given by

$$R_p(\alpha) = \frac{1}{2} \log\left(1 + \frac{(\sqrt{P_p} + a\sqrt{\alpha P_c})^2}{1 + a^2(1-\alpha)P_c}\right) \quad (25)$$
$$R_c(\alpha) = \frac{1}{2} \log(1 + (1-\alpha)P_c) \quad (26)$$

where $\alpha \in [0, 1]$ satisfies the condition $\frac{dR_c(\alpha)}{dR_p(\alpha)} \geq -1$.

Proof of Achievability: The random coding scheme described in the forward part of the proof of Theorem 3.1 (Section III-A) achieves the rates (25) and (26) stated in the theorem. We emphasize that, in this scheme, the primary receiver employs a single-user decoder.

Proof of Converse: See Appendix A.

The result of [24, Th. 3.5] further proves that the segment of the boundary of the capacity region complementary to the one stated above, i.e., the segment satisfying $\frac{dR_c}{dR_p} < -1$, is also given by (25) and (26). Thus, the entire capacity region of the low-interference IC-DMS is obtained.[8]

4) The Capacity of the Cognitive Radio Channel Under a Low-Interference Gain: The proof of Theorem 3.2 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 2.3 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 radio channel, we must solve for the positive root of the quadratic equation (18) in α. The solution is given by α^* in (14), hence the capacity is

$$R_c^* = \frac{1}{2} \log(1 + (1-\alpha^*)P_c). \quad (27)$$

Thus, we have established the proof of Theorem 3.1.

IV. The High-Interference-Gain Regime

The technique used to prove the converse of Theorem 3.2 also allows us to characterize the sum capacity of the $(1, a, b, 1)$-IC-DMS for any $a \geq 1$ and $b \in \mathbb{R}$, and the entire capacity region if a is sufficiently large and b is small enough. These two ancillary results are presented in this section.

A. 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 (25) and (26), i.e.,

$$C_{\text{sum}}(a) = \frac{1}{2} \log(1 + (\sqrt{P_p} + a\sqrt{P_c})^2). \quad (28)$$

Proof: See Appendix B.

B. The Benefit of Joint Code Design

We emphasize that the scheme that is optimal in the low-interference gain regime has the property that the primary receiver employs a single-user decoder. Contrary to this, we now observe that, when a is large enough, the optimal (jointly designed) IC-DMS code is such that the primary decoder d_p^n depends on the cognitive encoder e_c^n. First, we demonstrate an achievable scheme in the following lemma.

[8]The original version of our paper contained an upper bound for the $\frac{dR_c}{dR_p} < -1$ portion of the boundary as well, but was found to not be tight, in general. Thanks to H. El Gamal for pointing this out to us.

Lemma 4.1: Consider the $(1,a,b,1)$-IC-DMS. For every $\alpha \in [0,1]$, the rate pair (R_p, R_c) satisfying

$$R_p = \hat{R}_p(\alpha) \stackrel{\text{def}}{=} \frac{1}{2}\log(1+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2) \tag{29}$$

$$R_c = \hat{R}_c(\alpha) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1+\frac{(1-\alpha)P_c}{1+(b\sqrt{P_p}+\sqrt{\alpha P_c})^2}\right) \tag{30}$$

is achievable as long as

$$a \geq \frac{\sqrt{\alpha P_p P_c}}{K(\alpha)} + \sqrt{K(\alpha)+P_p(1+(b\sqrt{P_p}+\sqrt{\alpha P_c})^2)} \tag{31}$$

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

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 (29) and (30), 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 (31) stated in the theorem. □

In the achievable scheme described in Lemma 4.1, the primary receiver decodes the message of the cognitive user before decoding the message of the primary user. This scheme, in fact, is optimal for the $(1,a,b,1)$-IC-DMS when a is sufficiently large and b is small enough. This statement is made precise in the following theorem.

Theorem 4.2: A point (R_p,R_c) is on the boundary of the capacity region of the $(1,a,b,1)$-IC-DMS 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 (29) and (30), respectively;
2) a and b satisfy the condition given in (31);
3) $b \leq b_{\max}(\mu_\alpha, a)$ where $\mu_\alpha \stackrel{\text{def}}{=} -\frac{d^-\hat{R}_c(x)}{d\hat{R}_p(x)}|_{x=\alpha}$ and $b_{\max}(\mu,a)$ is defined in Appendix C.

Proof of Achievability: It is given in Lemma 4.1.

Proof of Converse: It is given in Appendix C.

As a gets larger and b gets smaller, Theorem 4.2 captures a larger portion of the boundary of the capacity region, until it finally characterizes the entire capacity region when $a \geq \sqrt{P_pP_c}/K(1)+\sqrt{K(1)+P_p(1+(b\sqrt{P_p}+\sqrt{P_c})^2)}$ and $b \leq b_{\max}(\mu_1,a)$.

V. 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.

A. Properties of the Optimal Scheme

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,[9] especially in the context of orthogonal frequency-division multiple access (OFDMA) [9], [10], where the primary users are allocated orthogonal time-frequency slots and the signal-to-interference-and-noise ratio (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.

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

[9]Classical request-to-send/clear-to-send (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.

statistics will cause no loss in the achievable rates.[10] Combined with the similar classical result for the standard AWGN channel [12], we see that the maximal rate expressed in Theorem 3.1 is achievable for all noise distributions.

B. Obtaining the Side Information

In practice, the cognitive radio is limited to obtaining the primary radio's codeword in a causal fashion—its acquisition thus introducing delays in the cognitive radio transmissions.[11] 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[12] 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 [15] 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.

In a practical implementation of a secondary system, the cognitive radio could be designed to efficiently exploit the automatic repeat request (ARQ) mechanism employed by the primary system. Most wireless systems today utilize ARQ protocols to increase the uplink packet decoding reliability: the retransmissions are either identical repeats of the erroneously decoded data frame (standard ARQ) or a subset of the parity bits of a punctured systematic error-correcting code (incremental redundancy hybrid-ARQ). In the presence of an ARQ protocol, the first transmission of the primary radio is usually underpowered (to minimize the energy consumption) and the primary base station is likely to send back a not-acknowledged signal (NAK). However, due to its proximity to the primary radio, the cognitive radio is very likely to have successfully decoded the data bits and could then communicate to its receiver during the next ARQ round(s) using the optimal code without the need to again listen and decode the primary transmission.

The key assumption here is that the cognitive radio is able to decode the ACK/NAK signals from the primary base station. Also, the cognitive radio must know the ARQ scheme being used (the particular code, puncturing pattern, and power increments in the case of IR Hybrid-ARQ). This information is typically periodically broadcast in the downlink of cellular systems for the benefit of new-coming primary users and a cognitive radio that is synchronized to the primary system could also extract this information.

C. 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 (p, f, g, c) cognitive radio channel with power constraints (P_p, P_c) and noise variances (N_p, N_s), as defined in Section II-A. However, the optimal cognitive encoder rule (15) must change slightly: the superposition scheme takes the form

$$X_c^n = \hat{X}_c^n + \frac{f^*}{|f|} e^{j\theta_p} \sqrt{\alpha \frac{P_c}{P_p}} X_p^n \tag{32}$$

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

$$\left(\frac{g}{c} + \frac{f^*}{|f|} e^{j\theta_p} \sqrt{\alpha \frac{P_c}{P_p}} \right) X_p^n \tag{33}$$

and the assumed noise is $\mathcal{CN}(0, N_s/|c|^2)$. The factor $e^{j\theta_p}$ in (32) essentially implements transmit beamforming to the primary receiver, hence ensuring that all the rates given by

$$0 \le R_p \le \log \left(1 + \frac{\left(|p| \sqrt{P_p} + |f| \sqrt{\alpha P_c} \right)^2}{N_p + |f|^2 (1-\alpha) P_c} \right) \tag{34}$$

$$0 \le R_c \le \log \left(1 + \frac{|c|^2 (1-\alpha) P_c}{N_s} \right) \tag{35}$$

are achieved in the underlying IC-DMS. As before, we can then choose $\alpha = \alpha^*$ [determined by (14)], so that $R_c^* = \log(1 + |c|^2 (1-\alpha^*) P_c / N_s)$ is achievable in the spirit of Definition 2.3 but with $R_p^* = \log(1 + |p|^2 P_p / N_p)$.

D. Communicating Without Channel–State Feedback From the Primary Base Station

In order to perform the complex baseband superposition coding scheme (32) and, implicitly, the Costa precoding for known interference (33), the cognitive radio must know each of the four parameters g, c, f, and p, both in magnitude and phase.

To obtain estimates for both the magnitude and phase of p and f, the cognitive radio would require additional help from the primary system in the form of channel–state feedback from the primary base station. In Section V-E, we present a method for accomplishing this, based on the assumption that the cognitive radio can extract the channel–state feedback intended for the primary radio. In this section, however, we present an alternative scheme which requires no feedback from the primary base station and which achieves the low-SNR capacity of a "fast-fading" cognitive radio channel under the assumption that the cognitive radio has no knowledge of p and f.

1) The Capacity in the Absence of Channel State Information: Suppose that, after having decoded X_p^n, the cognitive radio transmits the following n-symbol codeword:

$$X_c^n = \hat{X}_c^n + \sqrt{\alpha \frac{P_c}{P_p}} X_p^n \tag{36}$$

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

$$\left(\frac{g}{c} + \sqrt{\alpha \frac{P_c}{P_p}} \right) X_p^n \tag{37}$$

[10] 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.

[11] 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.

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

assuming the presence of $\mathcal{CN}(0, N_s/|c|^2)$ noise at the secondary base station. Note that this scheme does not require knowledge of the channel between the cognitive radio and the primary receiver, parameter f, or the channel between the primary radio and primary receiver, parameter p.[13] The cognitive radio does, however, require knowledge of the channel to its receiver, parameter c, and the channel between the primary radio and the secondary receiver, parameter g.

- *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.

This cognitive communication scheme is, in fact, optimal for the IC-DMS channel in which p and f are not known at the cognitive (and primary) radio, but are known at the primary receiver and are fixed for the duration of communication. In this case, it is straightforward to see that the boundary of the capacity region of this IC-DMS, parametrized by $\alpha \in [0,1]$, is given by

$$R_p = \log\left(1 + \frac{|p\sqrt{P_p} + f\sqrt{\alpha P_c}|^2}{N_p + |f|^2(1-\alpha)P_c}\right)$$
$$R_c = \log\left(1 + \frac{|c|^2(1-\alpha)P_c}{N_s}\right).$$

Observe that since $p, f \in \mathbb{C}$, the quantity $|p\sqrt{P_p} + f\sqrt{\alpha P_c}|^2$ could be arbitrarily small implying that the only way that the cognitive radio can ensure that it does not interfere with the primary radio is to completely shut off. In other words, the capacity, in the sense of Definition 2.4, of the cognitive radio channel is zero when p and f are not known at the cognitive radio and are fixed for the duration of communication.

However, such an extreme slow fading scenario is unlikely in practice: a more reasonable assumption might be that p and f are time varying (and still tracked at the primary receiver). Under the assumption of ideal interleaving, we can model these channel states $\{p[m]\}$ and $\{f[m]\}$ as ergodic random processes. In this case, the boundary of the capacity region of this "fast-fading" IC-DMS is given by

$$R_p = \mathbb{E}\left[\log\left(1 + \frac{|p\sqrt{P_p} + f\sqrt{\alpha P_c}|^2}{N_p + |f|^2(1-\alpha)P_c}\right)\right]$$
$$R_c = \log\left(1 + \frac{|c|^2(1-\alpha)P_c}{N_s}\right) \tag{38}$$

[13]Note that the parameter α in (36) could potentially depend on p or f. As it will turn out, its "optimal" value depends only on the *magnitude* of p, as will be shown in (42).

where the expectation is taken with respect to the stationary distribution of the ergodic processes. If we further assume that p and f are independent and their magnitude are deterministic, we can obtain the following low-SNR order approximation[14] for (38):

$$R_p = \Theta\left(\frac{|p|^2P_p + |f|^2\alpha P_c}{N_p + |f|^2(1-\alpha)P_c}\right) \text{ as } \max\left\{\frac{|p|^2P_p}{N_p}, \frac{|f|^2P_c}{N_p}\right\} \to 0. \tag{39}$$

On the other hand, the capacity R_p^* of the point-to-point channel from the primary radio to its receiver, in the absence of the cognitive radio, satisfies

$$R_p^* \stackrel{\text{def}}{=} \log\left(1 + \frac{|p|^2P_p}{N_p}\right)$$
$$= \Theta\left(\frac{|p|^2P_p}{N_p}\right), \qquad \text{as } \frac{|p|^2P_p}{N_p} \to 0. \tag{40}$$

Hence, in order to avoid causing interference to the primary user at low SNR, the following equation must be satisfied:

$$\frac{|p|^2P_p + |f|^2\alpha P_c}{N_p + |f|^2(1-\alpha)P_c} = \frac{|p|^2P_p}{N_p}. \tag{41}$$

If the cognitive radio tunes its parameter α such that

$$\alpha = \alpha_* \stackrel{\text{def}}{=} \frac{|p|^2P_p/N_p}{1 + |p|^2P_p/N_p} \tag{42}$$

this condition will be satisfied, hence $R_p = \Theta(\frac{|p|^2P_p}{N_p})$. The capacity of this "fast-fading" channel is then given by

$$R_c = \log\left(1 + \frac{|c|^2(1-\alpha_*)P_c}{N_s}\right)$$
$$\text{as } \max\left\{\frac{|p|^2P_p}{N_p}, \frac{|f|^2P_c}{N_p}\right\} \to 0. \tag{43}$$

In order to compute α_* given in (42), the cognitive radio only needs to know the received SNR of the primary transmission at the primary base station: $|p|^2P_p/N_p$. 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 $e^{R_p} - 1$. Hence, the assumption that the cognitive radio has knowledge of the received SNR of the primary transmission is implied by the assumption that the cognitive radio knows the (capacity achieving) codebook of the primary radio.

2) Achieving the Fast-Fading Capacity at Low SNR: The optimal cognitive communication scheme—both in the full-channel-state-information scenario, given by (32), and in the "fast-fading" scenario, given by (36)—requires that the cognitive radio's signal is perfectly time-synchronized with the primary radio's signal when it arrives at the primary base station, i.e.,

$$Y_p[m] = pX_p[m] + fX_p[m] + f\sqrt{\alpha\frac{P_p}{P_c}}\hat{X}_c[m] + Z_p[m].$$

[14]We say that $f(x) = \Theta(g(x))$ as $x \to 0$ if there exist constants $K_1, K_2 > 0$ such that $\lim_{x\to 0} \frac{f(x)}{g(x)} \le K_1$ and $\lim_{x\to 0} \frac{f(x)}{g(x)} \ge K_2$.

In practice, however, the delays from the two radios to the primary base station could be different,[15] resulting in a received signal given by

$$Y_p[m] = pX_p[m-l_p] + f\sqrt{\alpha\frac{P_c}{P_p}}X_p[m-l_c] + Z_{\text{total}}[m] \quad (44)$$

where $Z_{\text{total}}[m] = f\hat{X}_c[m-l_c] + Z_p[m]$ is the aggregate noise. Hence, to achieve perfect synchronization, the cognitive radio would have to advance its signal by $l_c - l_p$. Given that the performance of the primary system is at stake, the estimates for l_c and l_p would need to be as accurate as possible.

However, one can circumvent the sensitive and risky task of time-synchronization by exploiting the built-in multipath resilience of the primary receiver. Observe that (44) essentially describes a time-invariant two-tap multipath (or ISI) channel for the primary transmission. The primary system is naturally designed to tolerate a certain amount of delay spread[16] in the wireless channel. Transmit–receive architectures such as OFDM, or receivers such as the RAKE, in the case of direct-sequence spread spectrum, are two very common methods used to combat multipath fading in a wireless channel. Both of these schemes would yield[17] a power gain of $|p|^2P_p + |f|^2\alpha P_c$ at the primary base station (see, for instance, [21, Ch. 3], and references therein), thus achieving the rate (39) for the primary radio, at low SNR.

The received signal at the secondary base station, on the other hand, is given by

$$Y_p[m] = c\hat{X}_c[m-\tilde{l}_c] + \left(c\sqrt{\alpha\frac{P_c}{P_p}}X_p[m-\tilde{l}_c] + gX_p[m-\tilde{l}_p]\right) + Z_s[m]. \quad (45)$$

Hence, to achieve the "fast-fading" capacity promised in (43), $\hat{X}_c^n$ must be formed by Costa precoding for the interference term in parenthesis in (45). In order to do this, the cognitive radio must know the delays of the two radios with respect to the secondary base station: $\tilde{l}_c$ and $\tilde{l}_c$. In a stationary environment, these values can be accurately tracked by the secondary base station and fed back to the cognitive radio. Hence, the capacity of the fast-fading cognitive radio channel $R_c = \log(1+|c|^2(1-\alpha_*)P_c/N_s)$ is achieved.

3) Discreetly Entering the Primary Spectrum: Though the expression for α_* in (42) does not depend on $|f|$, we can see that (41) can approximately be satisfied even with $\alpha = 0$ when $|f|^2$ is very small. In other words, when $|f|$ is small, $\log(1 + |c|^2(1-\alpha_*)P_c/N_s)$ may be a very conservative estimate for the maximum achievable rate for the cognitive radio.

Since the cognitive radio has no information about $|f|$, a natural way for the cognitive radio to enter the spectrum of the primary system would be by first setting $\alpha = 0$ (transmitting purely its own information) and slowly *ramping* up its transmit power, from 0 to the maximum P_c, while listening for a NAK signal from the primary base station. If a NAK is detected, the cognitive radio should then slightly decrease its transmit power and continue monitoring the ARQ channel to ensure that its new transmit power level is acceptable.[18] This new power level, call it $P(k)$, should then be compared with the quantity $(1-\alpha_*)P_c$: if $P(k)$ is larger, it is kept as the new transmit power of the cognitive radio (with $\alpha = 0$). However, if $(1-\alpha_*)P_c > P(k)$, the cognitive radio could achieve a higher rate by using its maximum transmit power while still ensuring that the primary rate is unaffected, i.e., it should ramp up its power from $P(k)$ to P_c while increasing α from 0 to α_*.

E. Obtaining the Channel–State Information

In order to implement the optimal communication scheme of Costa coding and beamforming (32), the cognitive radio must obtain estimates of p and f from the primary base station. If the primary radio is, by default, able to obtain estimates of its channel to the primary base station (both magnitude and phase) via feedback from the primary base station, the cognitive radio could also potentially decode and use this information.[19] Under this assumption, we present the following simple algorithm that the cognitive radio can use to acquire the estimates of p and f.

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 uplink 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
$$X_c^n = \sqrt{\alpha\frac{P_c}{P_p}}X_p^n \quad (46)$$
where $\alpha \in [0,1]$.
3) The primary base station receives
$$\left(p + f\sqrt{\alpha\frac{P_c}{P_p}}\right)X_p^n + Z_p^n \quad (47)$$
and computes an estimate $\hat{h}$ of the overall channel gain $(p + f\sqrt{\alpha\frac{P_c}{P_p}})$ as it decodes m_p.
4) The quantized version of $\hat{h}$ is then broadcast on the control channel, in the usual way, 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 $f\sqrt{\alpha P_c/P_p}$ which is then multiplied by $\sqrt{\alpha P_p/P_c}$, to obtain an estimate for f.

[15]Even though the two radios are relatively close, their signal paths could easily fall into different taps in a very high bandwidth system.

[16]In other words, the channel from the primary radio to its receiver is frequency selective with multiple taps. For simplicity of notation, we are only displaying the tap corresponding to the most dominant path.

[17]The total delay spread between the primary and cognitive radio transmission must be less than the maximum tolerable delay spread in the primary system: $(l_c - l_p)/W < T_d \approx 10\mu s$.

[18]The NAK signal may be caused by interference from sources other than the cognitive radio. If no more NAKs are transmitted after the cognitive radio reduces its power, and if the power of the primary radio returns to its nominal value, the cognitive radio can be sure that it indeed was the culprit.

[19]The cognitive radio would have to be able to decode the dedicated control channel bits of the particular primary user in question.

Observe that a large α would cause a greater probability of a deep fade for the primary user since the quantity $|p + f\sqrt{\alpha P_c/P_p}|$, in step 3) above, would be more likely to be much smaller than $|p|$. 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 a NAK would be generated by the primary base station. The conservative solution is for the cognitive radio to use a small value of α and soft-combine several consecutive estimates of f, while simultaneously listening for the primary base-station NAK signals. As soon as the first NAK is detected, the α parameter should either be reduced or set to 0. In the case of a time-varying channel, the sequence of steps 1)–6) would need to be performed periodically.

Finally, note that the capacity of the cognitive radio channel R_c^* in Theorem 3.1 can only be achieved under the assumption of perfect channel–state estimation and feedback. In practice, the estimation and quantization errors will reduce the achievable rate and their effect can be modeled as additional independent Gaussian noise.

Appendix A
Proof of the Converse Part of Theorem 3.2

First, we observe that the rate region specified in Theorem 3.2 is a convex set in Proposition D.1. We will use the following standard result from convex analysis (see, for instance, [17]) 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. 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

$$Y_p^n = \frac{1}{a}X_p^n + X_c^n + \frac{1}{a}Z_p^n \tag{48}$$
$$Y_s^n = bX_p^n + X_c^n + Z_s^n. \tag{49}$$

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

Suppose that a rate pair (R_p, R_c) is achievable, in the sense of Definition 3.2, for the $(1,a,b,1)$-IC-DMS. Assuming that the messages (m_p, m_c) are chosen uniformly and independently, we have, 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 $\tilde{P}_{e,p}^{(n)} \to 0$ and $P_{e,s}^{(n)} \to 0$, respectively. We start with the following bound on nR_p:

$$\begin{aligned} nR_p &\overset{(a)}{=} H(m_p) \\ &= I\left(m_p; Y_p^n\right) + H\left(m_p|Y_p^n\right) \\ &\overset{(b)}{\leq} I(m_p; Y_p^n) + n\epsilon_{p,n} \\ &= h\left(Y_p^n\right) - h\left(Y_p^n|m_p\right) + n\epsilon_{p,n} \end{aligned} \tag{50}$$

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

$$\begin{aligned} nR_c &= H(m_c) \\ &= H(m_c) + H\left(m_c|Y_s^n, m_p\right) - H\left(m_c|Y_s^n, m_p\right) \\ &= I\left(m_c; Y_s^n|m_p\right) + H\left(m_c|Y_s^n, m_p\right) \\ &\overset{(a)}{\leq} I\left(m_c; Y_s^n|m_p\right) + n\epsilon_{s,n} \\ &= h\left(Y_s^n|m_p\right) - h\left(Y_s^n|m_p, m_c\right) + n\epsilon_{s,n} \\ &\overset{(b)}{\leq} h\left(Y_s^n|m_p\right) - h\left(Y_s^n|m_p, m_c, X_p^n, X_c^n\right) + n\epsilon_{s,n} \\ &\overset{(c)}{=} h\left(Y_s^n|m_p\right) - h\left(Z_s^n\right) + n\epsilon_{s,n} \end{aligned} \tag{51}$$

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 fact that Z_s^n is independent of (m_p, m_c) and hence also of (X_p^n, X_c^n).

Let $\tilde{Z}^n$ 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

$$\begin{aligned} h\left(\hat{Y}_p^n|m_p\right) &\overset{(a)}{=} h\left(\hat{Y}_p^n|m_p, X_p^n\right) \\ &\overset{(b)}{=} h\left(\hat{Y}_p^n - \frac{1}{a}X_p^n|m_p, X_p^n\right) \\ &= h\left(X_c^n + \frac{1}{a}Z_p^n|m_p, X_p^n\right) \\ &\overset{(c)}{=} h\left(X_c^n + Z_s^n + \tilde{Z}^n|m_p, X_p^n\right) \\ &\overset{(d)}{=} h\left(X_c^n + Z_s^n + \tilde{Z}^n|m_p\right) \\ &\overset{(e)}{=} h\left(\tilde{Y}^n + \tilde{Z}^n|m_p\right) \end{aligned} \tag{52}$$

where (a) and (d) hold since X_p^n is the output of a deterministic function[20] 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 $\tilde{Z}^n$, and (e) follows from the definition $\tilde{Y}^n \overset{\text{def}}{=} X_c^n + Z_s^n$. By similar reasoning, we can write

$$h\left(Y_s^n|m_p\right) = h(\tilde{Y}^n|m_p). \tag{53}$$

Combining the bounds in (50) and (51), we get

$$\begin{aligned} n(\mu R_p + R_c) &\leq \mu\left(h\left(Y_p^n\right) - h\left(Y_p^n|m_p\right)\right) \\ &\quad + h\left(Y_s^n|m_p\right) - h\left(Z_s^n\right) + \mu n\epsilon_{p,n} + n\epsilon_{s,n} \\ &\overset{(a)}{=} \mu h\left(Y_p^n\right) + h\left(Y_s^n|m_p\right) - \mu h\left(Y_p^n|m_p\right) \\ &\quad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n} \end{aligned}$$

[20] In general, the encoder E_p^n could be stochastic, i.e., depend on a random variable generated at the primary radio. In this case, inequalities (50) and (51) would need to include conditioning on E_p^n in addition to m_p [note that this can be done for free in (51) while it would yield a further upper bound in (50)]. Since the steps in the proof would be the same, we assume that the encoders are deterministic to simplify the notation.

$$\begin{aligned}&\stackrel{(b)}{=} \mu h\left(Y_p^n\right) + h(\tilde{Y}^n|m_p) - \mu h(\tilde{Y}^n + \tilde{Z}^n|m_p)\\&\quad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}\\&\stackrel{(c)}{\leq} \mu h(Y_p^n) + h(\tilde{Y}^n|m_p)\\&\quad - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}h(\tilde{Y}^n|m_p)} + e^{\frac{2}{n}h(\tilde{Z}^n)}\right)\\&\quad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n} \qquad (54)\end{aligned}$$

where (a) follows from the fact that $Z_s^n \sim \mathcal{N}(0, \mathbf{I}_n)$, (b) follows from equalities (52) and (53), and (c) follows from the conditional version of the entropy power inequality (see [22, Prop. I.1]).

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 jth component. We can upper bound $h(\tilde{Y}^n|m_p)$ as follows:

$$\begin{aligned}h(\tilde{Y}^n|m_p) &= h\left(\tilde{Y}^n|m_p, X_p^n\right)\\&\stackrel{(a)}{=} \sum_{j=1}^n h\left(\tilde{Y}_j|m_p, \tilde{Y}_1^{j-1}, X_{p,j}, X_{p,1}^{j-1}, X_{p,j+1}^n\right)\\&\stackrel{(b)}{\leq} \sum_{j=1}^n h(\tilde{Y}_j|X_{p,j})\\&\stackrel{(c)}{\leq} \sum_{j=1}^n \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}\left[\tilde{Y}_j^2\right] - \frac{\mathbb{E}[\tilde{Y}_j X_{p,j}]^2}{\mathbb{E}[X_{p,j}^2]}\right)\right)\\&\stackrel{(d)}{=} \sum_{j=1}^n \frac{1}{2}\log(2\pi e((1-\alpha_j)P_{c,j}+1)) \qquad (55)\\&\stackrel{(e)}{\leq} \frac{n}{2}\log(2\pi e((1-\alpha)P_c+1)) \qquad (56)\end{aligned}$$

where (a) follows from the chain rule and (b) follows from the fact that conditioning does not increase entropy, and (c) follows from [20, Lemma 1] (a more direct proof appears in [19, Lemma 1]). Equality (d) follows from the following argument. Since jointly Gaussian $X_{p,j}$, $\tilde{Y}_{p,j}$ achieve equality in (c) (by [20, Lemma 1]), we can without loss of generality let

$$X_{c,j} = \hat{X}_{c,j} + \sqrt{\alpha_j \frac{P_{c,j}}{P_{p,j}}} X_{p,j} \qquad (57)$$

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

$$\begin{aligned}P_{c,j} &\stackrel{\text{def}}{=} \frac{1}{2^{nR_c}}\sum_{j=1}^{2^{nR_c}} X_{c,j}^2\\P_{p,j} &\stackrel{\text{def}}{=} \frac{1}{2^{nR_p}}\sum_{j=1}^{2^{nR_p}} X_{p,j}^2. \qquad (58)\end{aligned}$$

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

$$\alpha P_c = \frac{1}{n}\sum_{j=1}^n \alpha_j P_{c,j} \qquad (59)$$

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 $h(Y_p)$ as follows:

$$\begin{aligned}h\left(Y_p^n\right) &\stackrel{(a)}{=} \sum_{j=1}^n h\left(Y_{p,j}|Y_{p,1}^{j-1}\right)\\&\stackrel{(b)}{\leq} \sum_{j=1}^n h(Y_{p,j})\\&\stackrel{(c)}{\leq} \sum_{j=1}^n \frac{1}{2}\log\left(2\pi e\mathbb{E}\left[Y_{p,j}^2\right]\right)\\&\stackrel{(d)}{=} \sum_{j=1}^n \frac{1}{2}\log\left(\frac{2\pi e}{a^2}(P_{p,j} + 2a\sqrt{\alpha_j P_{p,j} P_{c,j}} + a^2 P_{c,j} + 1)\right)\\&\stackrel{(e)}{\leq} \frac{n}{2}\log\left(\frac{2\pi e}{a^2}((\sqrt{P_p} + a\sqrt{\alpha P_c})^2 + a^2(1-\alpha)P_c + 1)\right)\end{aligned} \qquad (60)$$

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 (55), and (e) comes from Jensen's inequality applied to the $\log(\cdot)$ and the $\sqrt{\cdot}$ functions.

Let $f(x) \stackrel{\text{def}}{=} x - \frac{\mu n}{2}\log(e^{\frac{2}{n}x} + e^{\frac{2}{n}h(\tilde{Z}^n)})$ over $x \in \mathbb{R}$. Then, we can express the bound on our linear functional in (54) as

$$\begin{aligned}&n(\mu R_p + R_c)\\&\quad \leq \mu h\left(Y_p^n\right) + f(h(\tilde{Y}^n|m_p))\\&\qquad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n}. \qquad (61)\end{aligned}$$

Observe that as long as $\mu \leq 1$, $f(x)$ is increasing. Hence, we can obtain a further upper bound by substituting inequalities (56) and (60) into (61)

$$\begin{aligned}&n(\mu R_p + R_c)\\&\quad \leq \mu\frac{n}{2}\log\left(\frac{2\pi e}{a^2}((\sqrt{P_p} + a\sqrt{\alpha P_c})^2 + a^2(1-\alpha)P_c + 1)\right) \qquad (62)\\&\qquad + f\left(\frac{n}{2}\log(2\pi e((1-\alpha)P_c+1))\right)\\&\qquad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n} \qquad (63)\\&\quad \stackrel{(a)}{=} \mu\frac{n}{2}\log\left(\frac{2\pi e}{a^2}((\sqrt{P_p} + a\sqrt{\alpha P_c})^2 + a^2(1-\alpha)P_c + 1)\right) \qquad (64)\\&\qquad + \frac{n}{2}\log(2\pi e((1-\alpha)P_c+1))\\&\qquad - \mu\frac{n}{2}\log\left(2\pi e\left((1-\alpha)P_c + \frac{1}{a^2}\right)\right) \qquad (65)\\&\qquad - \frac{n}{2}\log(2\pi e) + \mu n\epsilon_{p,n} + n\epsilon_{s,n} \qquad (66)\end{aligned}$$

where (a) follows from the fact that

$$f(x) = x - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}x} + e^{\frac{2}{n}h(\tilde{Z}^n)}\right) \qquad (67)$$

$$= x - \frac{\mu n}{2}\log\left(e^{\frac{2}{n}x} + 2\pi e\left(\frac{1}{a^2}-1\right)\right) \quad (68)$$

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

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

$$\mu R_p + R_c \le \frac{\mu}{2}\log\left(1 + \frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right) + \frac{1}{2}\log(1+(1-\alpha)P_c). \quad (69)$$

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

$$\mu R_p + R_c \le \frac{\mu}{2}\log\left(1 + \frac{(\sqrt{P_p}+a\sqrt{\alpha_\mu P_c})^2}{1+a^2(1-\alpha_\mu)P_c}\right) + \frac{1}{2}\log(1+(1-\alpha_\mu)P_c). \quad (70)$$

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

B. *The $\mu \ge 1$ Case*

An outer bound to the boundary of the capacity region for the $\mu \ge 1$ case can be found in [11]. The original version of our paper contained this bound as a converse, but the bound was found to be not tight in general.

Appendix B
Proof of Corollary 4.1

The proof of this Corollary follows from Theorem 3.2 and Lemma D.1. In particular, we observe that the converse to Theorem 3.2 for $\mu \ge 1$ (see part B of Appendix A) holds for any $a > 0$ and $b \in \mathbb{R}$. However, from Lemma D.1 we see that the choice $\alpha = 1$ in (25) and (26) is optimal for any $a \ge 1$, as long as $\mu \ge 1$. Hence, the corollary is proved. $\square$.

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

Appendix C
Proof of the Converse Part of Theorem 4.2

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

$$Y_p = \mathbf{h}_p^T \mathbf{X} + Z_p \quad (71)$$
$$Y_s = \mathbf{h}_s^T \mathbf{X} + Z_s \quad (72)$$

where $\mathbf{h}_p = [1\, a]^T$ and $\mathbf{h}_s = [b\, 1]^T$. We choose $\mu \le 1$ in the linear functional $\mu R_p + R_c$ and recall that the optimal transmission vector $\mathbf{X}$ is Gaussian and given by

$$\mathbf{X} = X_{p1}\mathbf{u}_{p1} + X_{p2}\mathbf{u}_{p2} + X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2} \quad (73)$$

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 $(\mathbb{E}[\mathbf{X}\mathbf{X}^T])_{11} \le P_p$ and $(\mathbb{E}[\mathbf{X}\mathbf{X}^T])_{22} \le P_c$ on the achievable strategies in MIMO BC channel. We let

$$\mathbf{\Sigma}_p \stackrel{\text{def}}{=} \mathbf{u}_{p1}\mathbf{u}_{p1}^T + \mathbf{u}_{p2}\mathbf{u}_{p2}^T \quad (74)$$
$$\mathbf{\Sigma}_c \stackrel{\text{def}}{=} \mathbf{u}_{c1}\mathbf{u}_{c1}^T + \mathbf{u}_{c2}\mathbf{u}_{c2}^T \quad (75)$$

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

From [22], we know that the optimal encoding strategy is to generate X_p by Costa precoding[21] for $\mathbf{h}_p^T(X_{c1}\mathbf{u}_{c1} + X_{c2}\mathbf{u}_{c2})$. The rates achievable with such a scheme are

$$\begin{aligned} R_p &= \hat{R}_p(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \\ &\stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \mathbf{h}_p^T \mathbf{\Sigma}_p^* \mathbf{h}_p\right) \quad (76) \\ R_c &= \hat{R}_c(\mathbf{\Sigma}_p^*, \mathbf{\Sigma}_c^*) \\ &\stackrel{\text{def}}{=} \frac{1}{2}\log\left(1 + \frac{\mathbf{h}_s^T \mathbf{\Sigma}_c^* \mathbf{h}_s}{1 + \mathbf{h}_s^T \mathbf{\Sigma}_p^* \mathbf{h}_s}\right) \quad (77) \end{aligned}$$

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

$$\arg\max_{(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) \in \mathcal{S}(P_p, P_c)} \mu R_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) + R_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) \quad (78)$$

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

Since the per-antenna power constraints must be met with equality,[22] we can, without loss of generality, write

$$\mathbf{\Sigma}_p = \begin{bmatrix} \beta P_p & k_p \\ k_p & \alpha P_c \end{bmatrix}$$

where

$$k_p \in [-\sqrt{\alpha\beta P_p P_c}, \sqrt{\alpha\beta P_p P_c}] \quad (79)$$

$$\mathbf{\Sigma}_c = \begin{bmatrix} (1-\beta)P_p & k_c \\ k_c & (1-\alpha)P_c \end{bmatrix}$$

where

$$k_c \in [-\sqrt{\bar{\alpha}\bar{\beta} P_p P_c}, \sqrt{\bar{\alpha}\bar{\beta} P_p P_c}] \quad (80)$$

and $\beta \in [0,1]$ and $\alpha \in [0,1]$ and $\bar{\alpha} \stackrel{\text{def}}{=} 1-\alpha$ and $\bar{\beta} \stackrel{\text{def}}{=} 1-\beta$.

[21] 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.

[22] 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.

Substituting the covariance matrices (79) and (80) into (76) and (77), we get

$$\hat{R}_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \hat{R}_p(\beta, \alpha, k_p, a, b) \stackrel{\text{def}}{=} \frac{1}{2}\log(1+\beta P_p + 2ak_p + \alpha a^2 P_c) \tag{81}$$

$$\hat{R}_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \hat{R}_c(\beta, \alpha, k_p, a, b) \stackrel{\text{def}}{=} \frac{1}{2}\log\left(1+\frac{b^2(1-\beta)P_p + 2k_c b + (1-\alpha)P_c}{1+b^2\beta P_p + 2k_p b + \alpha P_c}\right). \tag{82}$$

The expression in (82) is maximized by choosing $k_c = \sqrt{(1-\beta)(1-\alpha)P_pP_c}$, 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

$$\hat{R}_p(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2}\log(1+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2) \tag{83}$$

$$\hat{R}_c(\mathbf{\Sigma}_p, \mathbf{\Sigma}_c) = \frac{1}{2}\log\left(1+\frac{(1-\alpha)P_c}{1+(b\sqrt{P_p}+\sqrt{\alpha P_c})^2}\right) \tag{84}$$

which exactly match the achievable rates given in Lemma 4.1. To this end, let $B(\mu, a)$ denote the set of all $b > 0$ such that the function

$$\max_{0\le\alpha\le 1} \mu\hat{R}_p(\beta,\alpha,k_p,a,b) + \hat{R}_c(\beta,\alpha,k_p,a,b) \tag{85}$$

is maximized, over all $\beta \in [0,1]$ and $k_p \in [-\sqrt{\beta\alpha P_pP_c}, \sqrt{\beta\alpha P_pP_c}]$, 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 $\mu \le 1$. Hence, we have proved the theorem.

APPENDIX D
SUPPORTING RESULTS

Proposition D.1: The rate region specified in Theorem 3.2 is a convex set.

Proof: A point $\mathbf{R} = (R_p, R_c)$ is in the rate region specified in Theorem 3.2 if and only if there exists $\alpha \in [0,1]$ such that

$$R_c \le \frac{1}{2}\log(1+(1-\alpha)P_c) \tag{86}$$

$$R_p + R_c \le \frac{1}{2}\log(1+a^2P_c + P_p + 2a\sqrt{\alpha P_p P_c}) + \frac{1}{2}\log\left(\frac{1+(1-\alpha)P_c}{1+a^2(1-\alpha)P_c}\right). \tag{87}$$

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 on the boundary of this region. Let $\alpha^{(1)} \in [0,1]$ and $\alpha^{(2)} \in [0,1]$ be their corresponding parameters in (86) and (87) for which the inequalities become tight. Then, for any $\lambda \in [0,1]$, we have that

$$\begin{aligned}\lambda R_c^{(1)} + (1-\lambda)R_c^{(2)} &\le \frac{\lambda}{2}\log\left(1+\left(1-\alpha^{(1)}\right)P_c\right) \\ &\quad + \frac{1-\lambda}{2}\log\left(1+\left(1-\alpha^{(2)}\right)P_c\right) &(88)\\ &\le \frac{1}{2}\log(1+(1-\alpha^*)P_c) &(89)\end{aligned}$$

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

$$\begin{aligned}&\lambda\left(R_p^{(1)}+R_c^{(1)}\right) + (1-\lambda)\left(R_p^{(2)}+R_c^{(2)}\right) \\ &\le \left[\frac{\lambda}{2}\log\left(1+a^2P_c+P_p+2a\sqrt{\alpha^{(1)}P_pP_c}\right) + \frac{1-\lambda}{2}\log\left(1+a^2P_c+P_p+2a\sqrt{\alpha^{(2)}P_pP_c}\right)\right] \\ &\quad + \left[\frac{\lambda}{2}\log\left(\frac{1+(1-\alpha^{(1)})P_c}{1+a^2(1-\alpha^{(1)})P_c}\right) + \frac{1-\lambda}{2}\log\left(\frac{1+(1-\alpha^{(2)})P_c}{1+a^2(1-\alpha^{(2)})P_c}\right)\right]. \qquad (90)\end{aligned}$$

We can further upper bound (90) as follows:

$$\begin{aligned}&\lambda\left(R_p^{(1)}+R_c^{(1)}\right) + (1-\lambda)\left(R_p^{(2)}+R_c^{(2)}\right) \\ &\stackrel{(a)}{\le} \frac{1}{2}\log\left(1+a^2P_c+P_p+2a\sqrt{P_pP_c}\left(\lambda\sqrt{\alpha^{(1)}}+(1-\lambda)\sqrt{\alpha^{(2)}}\right)\right) \\ &\quad + \frac{1}{2}\log\left(\frac{1+(1-\lambda\alpha^{(1)}-(1-\lambda)\alpha^{(2)})P_c}{1+a^2(1-\lambda\alpha^{(1)}-(1-\lambda)\alpha^{(2)})P_c}\right) \\ &\stackrel{(b)}{\le} \frac{1}{2}\log\left(1+a^2P_c+P_p+2a\sqrt{P_pP_c\alpha^*}\right) \\ &\quad + \frac{1}{2}\log\left(\frac{1+(1-\alpha^*)P_c}{1+a^2(1-\alpha^*)P_c}\right)\end{aligned}$$

where (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 $\log(\frac{1+(1-x)}{1+k(1-x)})$ (for constant $k \le 1$). Inequality (b) follows from Jensen's inequality applied to the square root function. This shows that $\lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$ is in the region as well, hence the region is a convex set. □

Lemma D.1:

$$\begin{aligned}\max_{0\le\alpha\le1}\frac{\mu}{2}\log\left(1+\frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right)+\frac{1}{2}\log(1+(1-\alpha)P_c) \\ = \frac{\mu}{2}\log(1+(\sqrt{P_p}+a\sqrt{P_c})^2) \qquad (91)\end{aligned}$$

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

$$
\begin{aligned}
&\max_{0\leq\alpha\leq 1}\frac{\mu}{2}\log\left(1+\frac{(\sqrt{P_p}+a\sqrt{\alpha P_c})^2}{1+a^2(1-\alpha)P_c}\right)+\frac{1}{2}\log(1+(1-\alpha)P_c) && (92)\\
&=\max_{0\leq\alpha\leq 1}\frac{1}{2}\log\left(\frac{(1+a^2(1-\alpha)P_c+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2)^{\mu}(1+(1-\alpha)P_c)}{(1+a^2(1-\alpha)P_c)^{\mu}}\right) && (93)\\
&\leq\max_{0\leq\alpha\leq 1}\frac{1}{2}\log\left(\frac{(1+a^2(1-\alpha)P_c+(\sqrt{P_p}+a\sqrt{\alpha P_c})^2)^{\mu}}{(1+a^2(1-\alpha)P_c)^{\mu-1}}\right) && (94)\\
&=\max_{0\leq\alpha\leq 1}\frac{1}{2}\log\left(\frac{(1+a^2P_c+P_p+2a\sqrt{\alpha P_p P_c})^{\mu}}{(1+a^2(1-\alpha)P_c)^{\mu-1}}\right) && (95)\\
&=\frac{\mu}{2}\log(1+(\sqrt{P_p}+a\sqrt{P_c})^2). && (96)
\end{aligned}
$$

Proof: On the one hand, we have (92)–(96) shown at the top of the page. On the other hand, the maximization problem in (91) can be lower bounded with $\frac{\mu}{2}\log(1+(\sqrt{P_p}+a\sqrt{P_c})^2)$, by choosing $\alpha=1$. Hence, the lemma is proved. □

ACKNOWLEDGMENT

The authors would like to thank H. El Gamal for pointing out an error in the original version of this paper, in the proof of converse of the IC-DMS in the original formulation of Theorem 3.2. The authors would also like to thank P. Mitran for his comments on the previous draft of this paper, G. Kramer for his suggestion that led to Section V-D1, and the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.
[2] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.
[3] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[4] H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 607–615, Sep. 1985.
[5] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.
[6] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.
[7] Federal Communications Commission Spectrum Policy Task Force, FCC, Report of the Spectrum Efficiency Working Group, Nov. 2002.
[8] *Federal Communications Commission, Cognitive Radio Technologies Proceeding (CRTP)*, ET Docket No. 03-108 [Online]. Available: http://www.fcc.gov/oet/cognitiveradio/
[9] *IEEE 802.16e Mobile WirelessMAN (R) Standard is Official*, IEEE Standards Association, Dec. 2005 [Online]. Available: http://standards.ieee.org/announcements/pr_p80216.html
[10] *Characteristics of the IEEE 802.16 Systems in the 2500–2690 MHz*, International Telecommunications Union draft, Dec. 2004 [Online]. Available: http://wirelessman.org/liaison/docs/L80216-04_42r2.pdf
[11] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," [Online]. Available: http://arxiv.org/abs/cs/0604107
[12] A. Lapidoth, "Nearest-neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.
[13] I. Maric, R. D. Yates, and G. Kramer, "The strong interference channel with unidirectional cooperation," in *Proc. Inf. Theory Appl. Inaugural Workshop*, La Jolla, CA, Feb. 2006.
[14] I. Maric, R. D. Yates, and G. Kramer, "The capacity region of the strong interference channel with common information," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2005.
[15] P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communications," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2041–2057, Jun. 2005.
[16] J. M. Peha and S. Panichpapiboon, "Real-time secondary markets for spectrum," *Telecommun. Policy*, vol. 28, pp. 603–618, Aug.–Sep. 2004.
[17] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1971.
[18] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.
[19] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1486–1495, Apr. 2005.
[20] J. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 711–716, Sep. 1987.
[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[22] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[23] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of Gaussian weak interference channels with degraded message sets," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Mar. 2006, pp. 1703–1708.
[24] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.

Aleksandar Jovičić received the B.S. degree in electrical engineering from the University of Wisconsin-Madison, Madison, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering (ECE) from the University of Illinois at Urbana-Champaign, Urbana, in 2003 and 2007, respectively.

Currently, he is a Staff Engineer at Qualcomm Flarion Technologies, Bridgewater, NJ.

Dr. Jovičić is the recipient of the 2007 Robert T. Chien Memorial Award for excellence in research from the Department of Electrical and Computer Engineering of the University of Illinois at Urbana-Champaign.

Pramod Viswanath (S'9–A'00–M'03) received the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley, in 2000.

Currently, he is an Associate Professor in the Electrical and Computer Engineering (ECE) Department, University of Illinois at Urbana-Champaign, Urbana. He was a Member of Technical Staff at Flarion Technologies until August 2001 before joining the ECE Department at the University of Illinois at Urbana-Champaign.

Prof. Viswanath is a recipient of the 2000 Eliahu Jury Award from the Electrical Engineering and Computer Science Department of the University of California, Berkeley, the 2000 Bernard Friedman Award from the Mathematics Department of the University of California, Berkeley, and the 2003 National Science Foundation (NSF) CAREER Award. He served as an Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2006–2008.