Encoding against an Interferer's Codebook

Ivana Marić, Nan Liu and Andrea Goldsmith
Stanford University, Stanford, CA
ivanam@wsl.stanford.edu, nanliu@stanford.edu, andrea@wsl.stanford.edu

***Abstract*—Motivated by cognitive radio applications, we consider mitigating the effect of interference by exploiting known properties about its signal structure. Specifically, we analyze communication between a source and destination with an interferer that induces random variations in the source-destination channel. The interferer transmits a sequence chosen uniformly from a randomly generated codebook, which has an i.i.d. structure, or a superposition structure. It is assumed that both the encoder and decoder know the interferer's codebook. We first provide a definition of capacity for these settings. When the encoder knows the interferer's message noncausally, it can use Gel'fand-Pinsker (GP) encoding to precode against interference. Alternatively, it can encode by taking into account that the interference is a codeword, to enable the decoder to decode both messages. It is demonstrated by an example that the latter can outperform GP encoding. Two upper bounds to the performance of this channel are then presented. Next, a more realistic scenario is considered in which the interference is learned at the cognitive encoder causally through a noisy channel. It is shown that for the case of i.i.d. generated interference, this information has no value. In contrast, when the interference is a codeword from an i.i.d. generated codebook, this fact can be exploited to obtain higher rates between the cognitive pair.**

I. Introduction

In wireless networks, source-destination pairs communicating simultaneously introduce interference to each other. This interference can be treated as noise, avoided by orthogonalizing transmissions, or handled by partial or full cancellation. Each of these approaches leads to a reduced performance; the interference can be fully cancelled without rate penalty only in the very strong interference regime [1]. Another way to manage interference opened up with the introduction of cognitive radio technology. Cognitive radios have the ability to obtain and exploit information about communications in their neighborhood. For example, this enables a cognitive encoder to decode messages transmitted in its vicinity. Information-theoretic approaches model these capabilities by assuming side information about the other user's message at the cognitive encoder. The simplest setting is shown in Fig. 1 where the (cognitive) transmitter-receiver pair communicates in the presence of a single interferer. The interfering codeword is assumed to be known non-causally without noise at the cognitive encoder. The scenario closely resembles the Gel'fand-Pinsker (GP) problem of coding for channels with random parameters (state) known at the transmitter [2]. The crucial difference is that in the latter setting, the state can be any typical realization of an i.i.d. process while in the former setting, the state is a codeword uniformly drawn from the interferer's codebook. Therefore, the number of possible states (i.e. codewords) depends on the interferer's rate and can be much smaller than the number of typical i.i.d. states. As analyzed in [3], this fact can be exploited to obtain higher rates between the cognitive pair.

In cognitive radio literature, it is often assumed that the interferer, i.e., the non-cognitive encoder, is willing to design its codebook jointly with the cognitive encoder (see [3], [4], [5], [6]). In reality, this is not always the case. In this work, we focus on the scenario where the interferer is using an i.i.d. generated codebook or a codebook randomly generated using superposition coding. An i.i.d. generated codebook is beneficial for the interferer to use if it is facing a single-user channel to its decoder, while a superposition codebook is beneficial for the interferer when it is performing rate-splitting or broadcasting information to two receivers of different channel quality.

For different realizations of codebooks used by the interferer, different performance can be achieved by the (cognitive) transmitter-receiver pair. Since the interferer's codebook is random, this performance is also random. In order to use one deterministic number to characterize the system performance, a new definition of capacity is needed. We first provide two operational definitions of capacity, and show that they are in fact equivalent, yielding the same number for capacity.

When the cognitive encoder knows the other message noncausally, it may use Gel'fand-Pinsker encoding to precode against interference, or it can take into account that the interference is a codeword, to enable its decoder to decode the interferer's codeword as well [3]. We show that the rates achievable with the scheme of [3] are achievable under the capacity definition presented in this paper. It is further demonstrated by an example that this scheme can outperform GP encoding. Two upper bounds on capacity are also presented.

Next, we consider a more realistic scenario, shown in Fig. 4, where the cognitive encoder, rather than obtaining the interferer's codeword noncausally without noise, learns symbols of the interferer's codeword causally through a noisy channel. For the two-sender, two-receiver cognitive radio channel, the causal scenario was considered in [4]. In this paper, it is shown that for the case of i.i.d. interference, this delayed information has no value, even if obtained without noise. In contrast, when the interference is a codeword, the encoder can potentially use this side information to decode the interferer's message and apply the same approach as in the noncausal case,

[1]This work by was supported in part from the DARPA ITMANET program under grant 1105741-1-TFIND, Stanford's Clean Slate Design for the Internet Program and the ARO under MURI award W911NF-05-1-0246.

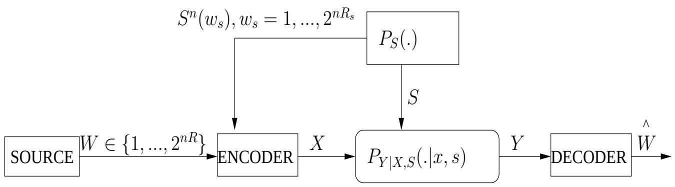

Fig. 1. Channel with an interfering codeword known at the transmitter.

for a fraction of time. In this paper, we demonstrate gains of this approach. In the next section we define the system model considered in this paper. Sections III and IV present our results for the scenarios in which the encoder knows the interference non-causally. The causal case is considered in Sec. V. This is followed by conclusions in Section VI. Due to space limitations, most of the proofs are omitted.

II. System Model

The discrete channel studied in this paper consists of two finite input alphabets $\mathcal{X}, \mathcal{S}$, a finite output alphabet $\mathcal{Y}$, and a probability distribution $p_{Y|XS}(y|x,s)$. The channel state S^n is not i.i.d., as in the GP problem. Instead, it takes the value of a row, in a uniform fashion, of a randomly generated codebook. In most of the paper, the randomly generated codebook is assumed to be i.i.d. generated with 2^{nR_s} rows and n columns, using $p_S(s)$. However, in Section IV, the randomly generated codebook takes on a superposition structure. In that case, the interferer rate-splits message $W_s = (W_{1s}, W_{2s})$ for some $W_{1s} \in \{1, \dots, 2^{nR_{1s}}\}, W_{2s} \in \{1, \dots, 2^{nR_{2s}}\}$ where $R_{1s} + R_{2s} = R_s$. More specifically, the encoding scheme at the interferer consists of:

- Choose a distribution $p(v)p(x_s|v)$.
- Generate $2^{nR_{1s}}$ codewords $v^n(w_{1s})$, $w_{1s} = 1, \dots, 2^{nR_{1s}}$ independently according to $P_V(\cdot)$.
- For each w_{1s}: Generate $2^{nR_{2s}}$ codewords $x_s^n(w_{1s}, w_{2s})$ using $\prod_{i=1}^n P_{X_s|V}(\cdot|v_i(w_{1s}))$, $w_{2s} = 1, \dots, 2^{nR_{2s}}$.

In the settings considered in the paper, an encoder wishes to send information to a decoder, in the presence of an interferer (see Fig. 1). Both the encoder and the decoder know the codebook of the interferer, $\mathcal{S}^n$. The channel is memoryless and time-invariant in the sense that

$$p(y_i|x^i, s^i, y^{i-1}) = p_{Y|XS}(y_i|x_i, s_i). \tag{1}$$

We will follow the convention of dropping subscripts of probability distributions if the arguments of the distributions are lower case versions of the corresponding random variables.

In Sections III and IV, we assume that at the beginning of a block, the encoder knows non-causally the codeword S^n, i.e., the interferer's message W_s, $\mathcal{W}_s = \{1, \dots, 2^{nR_s}\}$.

We define the capacity of such a communication scenario in two different ways.

Definition 1: A $(2^{nR}, n)$ code for the channel consists of: an index set $\{1, 2, \cdots, 2^{nR}\}$; encoding functions $f^n(\mathcal{S}^n)$: $\{1, 2, \cdots, 2^{nR}\} \times \{1, 2, \cdots, 2^{nR_s}\} \rightarrow \mathcal{X}^n$; and decoding functions: $g^n(\mathcal{S}^n) : \mathcal{Y}^n \rightarrow \{1, 2, \cdots, 2^{nR}\}$.

Note that while the encoding and decoding functions depend on the realization of the interferer's codebook, the size of the index set does not.

The probability of error is calculated as

$$P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) = \frac{1}{2^{nR} 2^{nR_s}} \sum_{W=w, W_s=w_s} \Pr[g^n(Y^n) \neq w | W = w, W_s = w_s]. \tag{2}$$

Rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes such that for any $\epsilon > 0$, for sufficiently large n, we have

$$\Pr\left[\mathcal{S}^n : P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) < \epsilon\right] > 1 - \epsilon. \tag{3}$$

The capacity of the channel, denoted as C_1, is defined as the supremum of all achievable rates.

Definition 2: A $(2^{nR(\mathcal{S}^n)}, n)$ code for the channel consists of: an index set $\{1, 2, \cdots, 2^{nR(\mathcal{S}^n)}\}$; encoding functions: $f^n(\mathcal{S}^n) : \{1, 2, \cdots, 2^{nR(\mathcal{S}^n)}\} \times \{1, 2, \cdots, 2^{nR_s}\} \rightarrow \mathcal{X}^n$; and decoding functions: $g^n(\mathcal{S}^n) : \mathcal{Y}^n \rightarrow \{1, 2, \cdots, 2^{nR(\mathcal{S}^n)}\}$.

Note that in this definition, not only the encoding and decoding functions, but also the size of the index set depend on the realization of the interferer's codebook, i.e., we allow the rate of the communication to change based on the realization of the interferer's codebook.

The probability of error is calculated as

$$P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) = \frac{1}{2^{nR(\mathcal{S}^n)} 2^{nR_s}} \sum_{W=w, W_s=w_s} \Pr[g^n(Y^n) \neq w | W = w, W_s = w_s]. \tag{4}$$

A rate $R(\mathcal{S}^n)$ is achievable if, for any $\epsilon > 0$, for sufficiently large n, there exists a sequence of codes $(2^{nR(\mathcal{S}^n)}, n)$ such that $P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) < \epsilon$. For each $\mathcal{S}^n$, let $C(\mathcal{S}^n)$ be the supremum of all achievable rates $R(\mathcal{S}^n)$. The capacity of the channel, denoted as C_2, is defined as

$$\sup_{\bar{R}} \left\{ \bar{R} : \Pr[\mathcal{S}^n : C(\mathcal{S}^n) > \bar{R} - \epsilon] > 1 - \epsilon \right\}. \tag{5}$$

We now explain the operational meaning of both definitions of capacity. It is clear that the interference degradation to the source-destination channel depends on the realization of the interferer's codebook.

In Definition 1, the rate of communication is the same for any realization of the interferer's codebook, however, the probability of error is not. For some realizations of the interferer's codebook, the communication rate R can be supported, and therefore, the average probability of error can be driven to zero. For others, rate R cannot be supported and, therefore, the average probability of error cannot be driven to zero. We want to find the supremum over all rates R that can be supported by the "majority" of possible interference codebooks, where majority here is in terms of a set with probability that is larger than $1 - \epsilon$.

Since the encoder knows which of the interferer's codebooks is in use, in Definition 2 we allow the encoder's communication rate to vary based on this known codebook. In fact, the encoder may adjust its rate up to the capacity of the channel, i.e., $C(\mathcal{S}^n)$. For different realizations of interference codebook, i.e., $\mathcal{S}^n$, the performance, i.e., $C(\mathcal{S}^n)$, is different. We take as capacity the rate that can be supported by the "majority" of possible interference codebooks, where majority here again is in terms of a set with probability that is larger than $1-\epsilon$.

In Section V, we take a more realistic approach as compared with Sections III and IV, and assume that the encoder, rather than knowing the interferer's codeword noncausally without noise, learns the interferer's codeword causally through an orthogonal, noisy DMC channel $p(z|s)$. The capacity definitions remain the same as in Definitions 1 and 2, with the exception that the encoding functions may only depend on message W and the previously received signals from the interferer, i.e., $(Z_1, Z_2, \cdots, Z_{i-1})$. More specifically, encoding at time i is

$$X_i = f_i(W, Z_1, \ldots, Z_{i-1}). \tag{6}$$

The problem studied in this paper is to characterize and bound the capacity of the (cognitive) transmitter-receiver pair as a function of the rate of the interferer's codebook, R_s.

The next theorem proves that the two capacity definitions are equivalent.

Theorem 1:

$$C_1 = C_2 \triangleq C. \tag{7}$$

Proof: The proof is given in Appendix A. ■

From (7) we see that C is the appropriate capacity metric for this system.

The following lemma shows that if a code of rate R is good on average, i.e., that the error probability of the code averaged over all interferer's codebooks can be made arbitrarily small, then R is achievable by Definition 1.

Lemma 1: Consider a $(2^{nR}, R)$ code that satisfies

$$E[P_e(S^n)] \le \epsilon$$

as $n \to \infty$, for some $\epsilon > 0$. Then, R satisfies (3) and is thus achievable under Definition 1.

In the rest of the paper, we use Definition 1 and Lemma 1 to lower bound the capacity, and Definition 2 to upper bound the capacity.

III. Results for the GP Problem with the Codebook

A. Achievable Rates

We restate the achievability result from [3, Lemma 1]. Using Lemma 1, it will then be easy to show that this rate is achievable also under the capacity definition of Sec. II.

Lemma 2: For the channel of Fig. 1, the rate

$$R \le \max_{p(u|s), f(\cdot)} \min\{I(X;Y|S), \max\{I(U;Y) - I(U;S), I(X,S;Y) - R_s\}\} \tag{8}$$

is achievable. For $I(S;U,Y) \le R_s \le H(S)$, binning achieves the GP rate given by $I(U;Y) - I(U;S)$ in (8). For $R_s \le I(S;U,Y)$, superposition coding achieves the rate given by the minimization term in (8).

Proof: It was shown in [3] that for the rate (8), the average error probability can be made arbitrarily small. Lemma 1 then implies that rate (8) is also achievable under Definition 1. ■

Remark 1: For the distribution $p(u|s), f(\cdot)$ of Lemma 2, it holds that $I(X,S;Y) = I(U,S;Y)$.

In Lemma 2, when the rate of the interferer's codebook is small, i.e., $R_s \le I(S;U,Y)$, it does not place too much burden on the system to decode W_s. When R_s is large, treating it as an i.i.d. sequence and using the GP scheme is better than requiring the receiver to decode W_s.

B. Example

We next provide an example to demonstrate that taking into account that the state is a codebook helps, i.e., we show that the achievable rate of Lemma 2 can improve the GP rate, unlike in the case of i.i.d. S^n sequences. We consider the channel that models computer memory with random defects and noise, first proposed by Kuznetsov and Tsybakov and analyzed in [2].

The channel has a binary input and a binary output, i.e., $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and the channel state takes three possible values, $\mathcal{S} = \{0,1,2\}$. The channel is specified by three parameters, λ, p, q which are all between 0 and $1/2$. The channel transition probability $p(y|x,s)$ is given by

$$p(y|x, s=0) = \begin{cases} 1-q & \text{if } (x,y) = (0,0) \text{ or } (x,y) = (1,0) \\ q & \text{otherwise} \end{cases}$$

$$p(y|x, s=1) = \begin{cases} q & \text{if } (x,y) = (0,0) \text{ or } (x,y) = (1,0) \\ 1-q & \text{otherwise} \end{cases}$$

$$p(y|x, s=2) = \begin{cases} p & \text{if } (x,y) = (0,1) \text{ or } (x,y) = (1,0) \\ 1-p & \text{otherwise} \end{cases}$$

In other words, in states $s=0$ and $s=1$, the channel output does not depend on the channel input, and in state $s=2$, the channel is a binary symmetric channel with crossover probability p. The distribution of the state S is

$$p(s) = \begin{cases} \lambda & \text{if } s=0 \text{ or } s=1 \\ 1-2\lambda & \text{otherwise} \end{cases} \tag{9}$$

The rate achievable in (8) may be written as

$$R \le \max_{p(u|s), f(\cdot)} \max\{\min\{I(X;Y|S), I(X,S;Y) - R_s\}, I(U;Y) - I(U;S)\} \tag{10}$$

since the GP rate $I(U;Y) - I(U;S)$ is always less than the rate achievable where the receiver also knows the realization of the channel state perfectly, i.e., $I(X;Y|S)$.

From (10), to find the achievable rate, all we need to characterize are

$$\max_{p(x|s)} \min\{I(X;Y|S), I(X,S;Y) - R_s\} \tag{11}$$

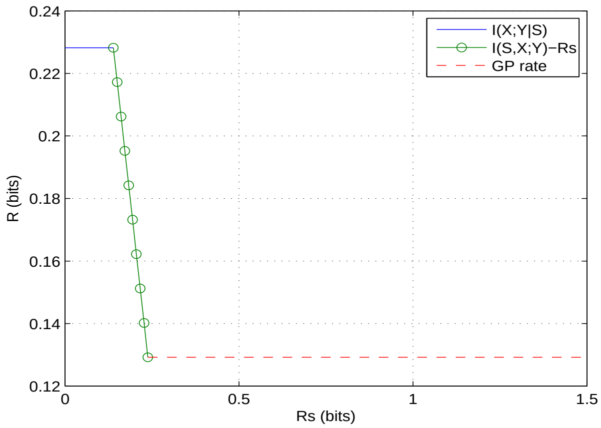

Fig. 2. An achievable rate in the example with $\lambda = 1/4, p = 1/8, q = 1/5$. For smaller values of R_s, decoding the interferer's message achieves a higher rate than GP encoding.

and

$$\max_{p(u|s), f(\cdot)} I(U;Y) - I(U;S). \quad (12)$$

The term in (12) is the GP rate given by [2, Proposition 4]. Based on the symmetry of the channel, it is straightforward to show that the optimal $p(x|s)$ that maximizes the term in (11) is $p(x|s=2)$ being Bernoulli $1/2$ and $p(x|s=0,1)$ being Bernoulli with any arbitrary parameter.

We plot the achievable rate, i.e., the maximum of (11) and (12) in Fig. 2. The parameters of the channel are $\lambda = 1/4, p = 1/8, q = 1/5$. We observe from Fig. 2 that in this example, even though S^n is i.i.d. generated, treating it as i.i.d. and using GP encoding is not optimal when the rate R_s is small. Rather, it is better for the receiver to fully decode the channel state.

Remark 2: For some channels, such as the example in this section with $q = 0$ [7], and the Gaussian channel with state [8], the GP rate is equal to $\max_{p(x|s)} I(X;Y|S)$. In such channels, S^n belonging to an i.i.d. generated codebook with rate $R_s < H(S)$ does not offer any performance gains, as compared to the case where the state S^n is i.i.d.

C. Converse

In this subsection, we provide two upper bounds on the capacity. First, we prove a tight converse in the case where $R_s \le \min_{p(x|s)} I(S;Y)$.

Lemma 3: The achievable rate has to satisfy

$$R \le \max_{p(x|s)} I(X;Y|S) \quad (13)$$

with $p(s,x,y) = p(s)p(x|s)p(y|x,s)$.

The upper bound provided in Lemma 3 is the same as the capacity of a system where S^n is i.i.d. with $p(s)$, and the receiver, as well as the transmitter, knows S^n.

Remark 3: We note that when $R_s \le \min_{p(x|s)} I(S;Y)$, the achievable rate given by (8) is equal to $\max_{p(x|s)} I(X;Y|S)$, the same as the converse in (13), yielding the capacity.

To obtain a general upper bound, we modify the GP converse to take into account that the states are constrained to be one of the interferer's codewords. As in GP case, we bound the rate by Fano's inequality as

$$\begin{aligned} nR(\mathcal{S}^n) &\le I(W;Y^n) \\ &= \sum_{i=1}^{n} I(W, S_{i+1}^n; Y_i|Y^{i-1})] - I(S_i; Y^{i-1}|W, S_{i+1}^n)] \\ &= \sum_{i=1}^{n} [H(Y_i|Y^{i-1}) - H(Y_i|U_i)] \\ &\quad - [H(S_i|S_{i+1}^n) - H(S_i|U_i)] \end{aligned} \quad (14)$$

where we defined $U_i = [W S_{i+1}^n Y^{i-1}]$. As in [9], we next let

$$\Delta_S^n = \sum_{i=1}^{n} H(S_i) - H(S^n) = \sum_{i=1}^{n} H(S_i) - nR_s \quad (15)$$

and similarly

$$\Delta_Y^n = \sum_{i=1}^{n} H(Y_i) - H(Y^n)$$

and use it in (14) to obtain

$$\begin{aligned} nR(\mathcal{S}^n) &\le \sum_{i=1}^{n} [H(Y_i) - H(Y_i|U_i)] - [H(S_i) - H(S_i|U_i)] \\ &\quad + \Delta_S^n - \Delta_Y^n \\ &\le n \max_i [H(Y_i) - H(Y_i|U_i)] - [H(S_i) - H(S_i|U_i)] \\ &\quad + \Delta_S^n - \Delta_Y^n \\ &\le nR^* + \Delta_S^n - \Delta_Y^n \end{aligned} \quad (16)$$

where R^* denotes the GP rate in the case when S^n is i.i.d. Since any $R(\mathcal{S}^n)$ satisfies (16), we conclude that $C(\mathcal{S}^n)$ is also bounded as in (16). Following the definition of the capacity C_2, we obtain an upper bound

$$C \le R^* + \frac{1}{n} \max_{\mathcal{S}^n} \{\Delta_S^n - \Delta_Y^n\}. \quad (17)$$

Remark 4: The obtained upper bound gives only an n-letter characterization and is thus not computable. The bound can further be loosened to obtain a single-letter characterization. One way to do so, for example, is to use the fact that $\Delta_Y^n \ge 0$ and from bound Δ_S^n in (15) as

$$\frac{\Delta_S^n}{n} \le \max_i H(S_i) - R_s \quad (18)$$

We obtain

$$C \le R^* + \max_{p(s)} H(S) - R_s. \quad (19)$$

Then, evaluation of the tightness of this or similar bounds can be performed.

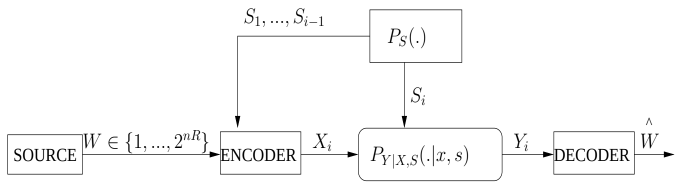

Fig. 3. Channel with random states known with delay at the transmitter.

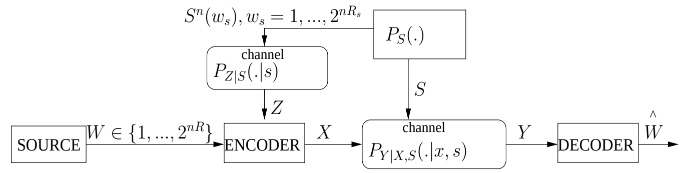

Fig. 4. Channel with an interfering codeword observed through a noisy channel at the cognitive encoder.

IV. Improvement when the Interferer Uses a Superposition Code

In this section, we examine the performance when the code of the interferer in Fig. 1 is more 'structured'. Hence, rather than assuming the codebook is i.i.d. generated, we assume that it is randomly generated by superposition of two codebooks with rates R_{1s}, R_{2s}, and $R_{1s} + R_{2s} = R_s$, defined in Sec. II.

Lemma 4: For the encoder and the decoder of Fig. 1, the rate R is achievable if it satisfies

$$R \leq I(U; Y|V) - I(U; X_s|V) \tag{20}$$

$$R + R_{1s} \leq I(V, U; Y) - I(U; X_s|V) \tag{21}$$

for a joint distribution that factors as $p(u|v)p(y|x, x_s)$, and x is chosen to be a deterministic function of (v, x_s, u). R_{1s} and $p(v)p(x_s|v)$ are determined by the interferer.

Remark 5: Rates of Lemma 2 are a special case of (20)-(21): for $V = \emptyset$ in (20)-(21), the GP rate given by the term $I(U; Y) - I(U; S)$ in (8) is obtained. In this case, the cognitive encoder does not partially decode W_s. We have the opposite scenario for $X_s = \emptyset$: the cognitive encoder fully decodes W_s and we obtain the minimization term in (8). Thus, rates (8) are obtained for two different choices of V.

V. Results for GP Problem with the Codebook, Delay and Noise

We next consider a more realistic scenario where the cognitive encoder learns the interferer's sequence causally through an orthogonal, noisy DMC channel, $p(z|s)$. Encoding at time i can only depend on message W and past received symbols $(Z_1, Z_2, \cdots, Z_{i-1})$. Similarly to the definition of the relay channel [10], the encoder cannot instantaneously use the knowledge of Z_i at time i. In Sec. V-A, we study the case where the interfering sequence is i.i.d. with $p(s)$. In Sec. V-B, we focus on the case where the interferer's sequence is uniformly chosen from an i.i.d. generated codebook with $p(s)$.

A. I.I.D. Interference Sequence with Delay and Noise

We first consider a noiseless channel from S to Z, i.e., $Z = S$, as shown in Fig. 3. The obtained results serve as an upper bound on the case of the noisy channel.

For channels with random states, non-causal and causal but instantaneous knowledge of the channel state can be exploited for higher rates [2], [11]. But what happens if the random state is learned noiselessly, causally, but not instantaneously? More specifically, we consider an encoding function:

$$X_i = f_i(W, S_1, \ldots, S_{i-1}) \tag{22}$$

rather than the channel with causal and instantaneous knowledge, $X_i = f_i(W, S_1, \ldots, S_i)$, considered by Shannon [11]. For a discrete memoryless channel and i.i.d. states, this knowledge cannot help, as formalized by the following lemma.

Lemma 5: For the DMC channel with a delayed knowledge of an i.i.d. state shown in Fig. 3, the capacity is given by

$$R = \max_{p(x)} I(X; Y). \tag{23}$$

Proof: The proof is given in Appendix B. ■

Lemma 5 implies that the rate equals the rate achieved with no side information at the encoder.

When the channel from S to Z is noisy, the capacity of the (cognitive) transmitter-receiver pair is upper bounded by the result of Lemma 5, and hence we may conclude that, again, receiving the past Z_is does not increase capacity and it is optimal to ignore them.

B. Interferer Codeword with Delay and Noise

We next focus on the case where the interferer's sequence comes from a codebook. In the cognitive network scenarios, a state is a noisy observation of a symbol from the interferer's codeword (see Fig. 4). For that reason, observing a sequence of states may allow the cognitive radio to decode message W_s. Delay introduced at the cognitive encoder when decoding W_s will drastically impact the rate performance, as discussed next.

First, we show that the cognitive encoder should not wait until the end of the block to decode W_s. Denote the interferer's message sent in block $b-1$ as $W_s(b-1)$. Consider the scenario in which the cognitive encoder is able to decode $W_s(b-1)$ at the end of block $b-1$. It can then use $W_s(b-1)$ for encoding in the next block:

$$X^n(b) = f_b(W(b), W_s(1), \ldots, W_s(b-1)). \tag{24}$$

Similar encoding is, for example, encountered at a relay performing the decode-and-forward cooperative strategy [12]. Since messages $W_s(1), \ldots, W_s(b-1)$ sent in different blocks are i.i.d, encoding (24) can be cast into encoding given by (22), by viewing each block b as a one channel use i in (22).

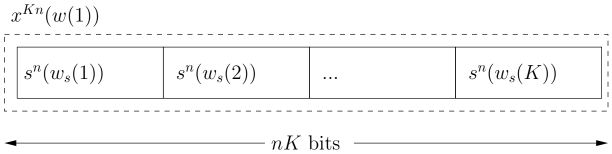

Fig. 5. Encoding over K blocks at the cognitive encoder.

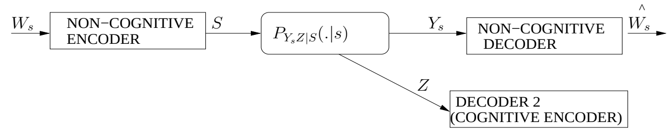

Fig. 6. Channel from the interferer to its receiver and to the cognitive encoder (labeled decoder 2).

Then, Lemma 5 implies that knowing W_s of the previous block cannot improve the rate. This is in sharp contrast to the case without delay discussed in Sec. III-A, where GP encoding can improve the performance. More specifically, Lemma 5 implies

$$nR \le \max_{p(x^n)} I(X^n;Y^n). \tag{25}$$

However, this rate can be achieved even if

$$X^n(b) = f_b(W(b)). \tag{26}$$

To achieve (26), the encoder uses a random code $(2^{nKR}, R)$, as shown in Fig. 5. Thus, the encoder encodes over K blocks of length n. As $K \to \infty$, the average error probability becomes small and (25) is achieved. Thus, knowing the interferer's message sent in the previous block, does not bring rate gains.

In general, since the encoding in Fig. 4 is given by (6), it is not necessary for the cognitive encoder to decode after receiving the whole block Z^n. If decoding of W_s can be done within a block, delayed employment of the encoding schemes of Sec. III-A is possible. Next, we present a simple encoding scheme that exploits delayed knowledge of the interferer's codeword. A similar approach was considered for the two-sender, two-receiver cognitive radio network in [4]. The achievable scheme applies to the case when the channel from the interferer to the cognitive encoder is better than the channel to its receiver. Then, the cognitive encoder can decode W_s *before* the intended receiver and have non-causal knowledge of the interferer's codeword for the rest of the block. This scheme depends on the rateless property of the random code, formalized next.

Lemma 6: Consider a transmitter and a receiver communicating in the presence of another receiver, as shown in Fig. 6. The discrete, memoryless channel is described by $p(y_s, z|s)$. The communicating pair is oblivious of the second receiver and communicates at rate

$$R_s = \max_{p(s)} I(S;Y_s). \tag{27}$$

Assume that there exists $n_1 < n$ such that

$$nR_s \le n_1 I(S;Z) \tag{28}$$

Then, decoder 2 can decode W_s with arbitrarily small error probability after n_1 channel uses.

We use (28) to define α as the fraction of symbols after which decoder 2 can decode W_s:

$$\alpha = \frac{n_1}{n} = \frac{R_s}{I(S;Z)}. \tag{29}$$

Before we proceed with the main result of this subsection, we prove a lemma which provides an achievable rate when the encoder is not cognitive (see Fig. 7), i.e., it does not know anything about W_s, though it knows the realization of the i.i.d. codebook $\mathcal{S}^n$. Related scenarios have been studied in [13].

Lemma 7: For the channel of Fig. 7, the rate

$$R \le \max_{p(x)} \max\{\min\{I(X;Y|S), I(X,S;Y) - R_s\}, I(X,Y)\} \tag{30}$$

is achievable.

Lemma 7 states that when the transmitter does not know the interferer's i.i.d. generated codeword, it can either perform like a transmitter in a multiple access channel, and allow the receiver to decode both W and W_s, or it could transmit at a rate which allows the receiver to treat S^n as i.i.d. noise.

Applying the rateless property of Lemma 6 to the cognitive encoder, we propose the following achievable scheme: the encoder splits its message W into two independent messages, W_1 and W_2. For the first n_1 symbols, it sends W_1 as if it does not know anything about the channel states. In this case, the rate of the state codebook is $nR_s/n_1 = R_s/\alpha$. From (30), the rate achievable is

$$\begin{aligned} R_1 &= \max\left\{ \max_{p(x)} \min\left\{ I(X;Y|S), I(S,X;Y) - \frac{1}{\alpha}R_s \right\}, \right. \\ &\qquad \left. \max_{p(x)} I(X;Y) \right\} \\ &= \max\left\{ R_{11}, R_{12} \right\} \end{aligned} \tag{31}$$

where we denote the first and the second term in the maximization as R_{11} and as R_{12}, respectively.

Based on Lemma 6, the transmitter is able to decode the state sequence after the first n_1 symbols. Thus, starting from symbol $n_1 + 1$, the transmitter knows $S^n_{n_1+1}$. If $R_{11} \ge R_{12}$, then, after n_1 symbols, not only the transmitter, but also the receiver knows the codeword of the interferer, i.e., W_s. Thus, for the remaining $n - n_1$ symbols, the rate achievable is

$$R_0 = \max_{p(x|s)} I(X;Y|S) \tag{32}$$

On the other hand, if $R_{12} > R_{11}$, then only the transmitter knows W_s, or in other words, $S^n_{n_1+1}$, and from symbol n_1+1, the problem becomes the same as in Sec. III, since the state $S^n_{n_1+1}$ is uniformly picked from an i.i.d. generated codebook. Hence, from symbol $n_1 + 1$ to n, we use the encoding scheme of Sec. III-A, and achieve the rate of Lemma 2. Note that the

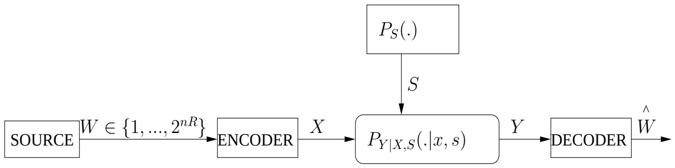

Fig. 7. Channel with an interfering codeword not known at the transmitter.

block length is $n - n_1$, and the rate of the state codebook is $\frac{n}{n-n_1} R_s = \frac{1}{1-\alpha} R_s$. The rate achievable is

$$R_2 = \max_{p(u|s),p(x|u,s)} \max \left\{ \min \left\{ I(X;Y|S), I(X,S;Y) - \frac{1}{1-\alpha} R_s \right\}, I(U;Y) - I(U;S) \right\}. \quad (33)$$

Hence, the rate achievable by this two stage scheme is

$$R_3 = \begin{cases} \alpha R_{11} + (1-\alpha) R_0 & \text{if } R_{11} \geq R_{12} \\ \alpha R_{12} + (1-\alpha) R_2 & \text{if } R_{11} < R_{12} \end{cases}. \quad (34)$$

On the other hand, one simple achievability scheme is for transmitter 1 not to use the delayed knowledge about the state sequence. In this case, the rate achievable by Lemma 7 is (30), which is denoted as R_4. Hence, the achievable rate with delayed knowledge of interferer's codeword is $R_{\text{delay}} = \max\{R_3, R_4\}$.

We illustrate the rate improvement for the transmitter with delayed knowledge, over the transmitter with no knowledge, with an example in Fig. 8. We assume a Gaussian channel from the transmitter to the receiver in the presence of interference S as

$$Y = X + S + Z_1 \quad (35)$$

and the channel from the interferer to the receiver of the cognitive encoder as

$$Z = S + Z_2 \quad (36)$$

where Z_1 and Z_2 are Gaussian with zero mean and variances $N_1 = 1$ and $N_2 = 0.5$, respectively. The power constraint of the transmitter is $P = 5$ and the distribution for generating the i.i.d. codebook at the interferer is zero-mean Gaussian with variance $Q = 3$. The star line illustrates the capacity in the case where the transmitter has no knowledge of the channel state, i.e., R_4 [13]. The solid line illustrates the achievable rate of our two-stage proposed scheme, i.e., R_3. By taking the larger of R_3 and R_4, we obtain the performance when the transmitter has a delayed noisy version of the interference, i.e., R_{delay}, illustrated by the circle line. As can be seen from the figure, in contrast to the case when the state is i.i.d., when the state is a codeword from an i.i.d. generated codebook, delayed knowledge does improve the performance of the system.

Finally, we illustrate the difference in the performance for the case of delayed knowledge of the interferer's signal and the case of the noncausal knowledge of the interferer's signal at the cognitive encoder, studied in Sec. III. We use the

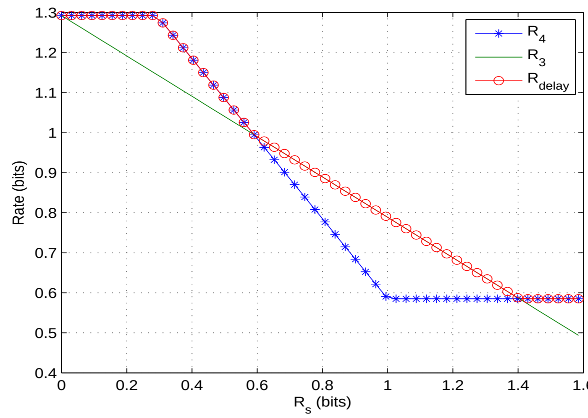

Fig. 8. Performance when the cognitive encoder learns the interferer's sequence causally with noise in the Gaussian channel with a Gaussian codebook at the interferer.

example of Sec. III-B, where in the delayed case, we assume the interferer's signal is received noiselessly at the cognitive encoder, i.e., $Z = S$. The comparison is shown in Fig. 9. The solid line shows the performance of noncausal knowledge of the interferer's signal at the cognitive encoder, i.e., the rates of Lemma 2, while the dash line shows the performance with the delayed knowledge of the interferer's signal at the cognitive encoder, i.e., R_{delay}. In this particular example, when the interferer's rate, R_s, is small enough for the receiver to decode, it makes no difference whether the transmitter knows the interferer's signal. However, when R_s is large, noncausal knowledge of the interferer's signal provides the Gel'fand-Pinsker rate while delayed knowledge of the interferer's signal provides the GP rate only a fraction of the time, i.e., the last $n - n_1$ symbols.

VI. Conclusion

Motivated by cognitive radio networks, a problem of communication over a channel with random parameters is formulated. The randomness is introduced by a nearby interferer communicating to its receiver. In the classical setting, the state is any typical realization of an i.i.d. process causally or noncausally known at the encoder. In contrast, the distinct features of the considered problem are that the state of the channel is a codeword uniformly drawn from the interferer's codebook, and that the state knowledge may be available after an additional delay. We initially assume noncausal knowledge of the state (interferer's codeword) at the cognitive encoder. We present an example demonstrating that there are benefits from taking into account that the interference is a codeword, and give two upper bounds to the performance. The impact of the delay in learning the state at the cognitive encoder is then investigated. In particular, we consider a scenario where the cognitive encoder learns the interferer's codeword causally through a noisy channel. While in the case of i.i.d. interference this information has no value, we present a simple encoding scheme for the cognitive radio settings that brings rate gains.

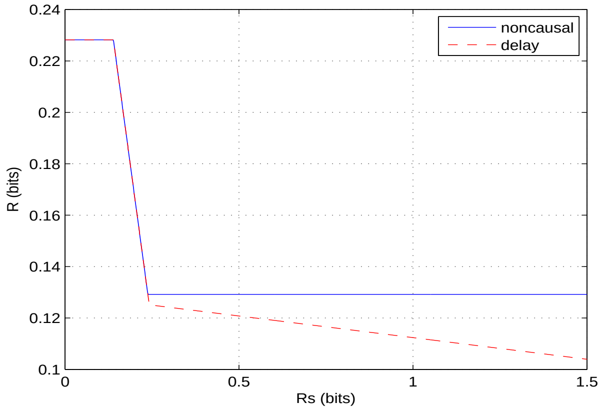

Fig. 9. Rate comparison for cases of the delayed and noncausal knowledge of the interferer's signal at the cognitive encoder.

We illustrate the impact of delay with two examples.

Appendix A
Proof of Theorem 1

We will first show that $C_1 \leq C_2$. To show this, it is equivalent to show that any rate, R, achievable by the first definition is a valid candidate for $\bar{R}$, see (5), in the second definition. Then, by taking the supremum of both R and $\bar{R}$, we get $C_1 \leq C_2$. Fix an $\epsilon > 0$. Let R be achievable in the first definition, i.e., there exists a sequence of $\left(2^{nR}, n\right)$ codes such that for the fixed ϵ and sufficiently large n, we have (3). Let the set of $\mathcal{S}^n$ that satisfies $P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) < \epsilon$ for this sequence of codes be called $\mathcal{A}$. For $\mathcal{S}^n \in \mathcal{A}$, because for this sequence of codes $P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n)) < \epsilon$, in Definition 2, R is an achievable rate for $\mathcal{S}^n$, i.e., by Definition 2, $C(\mathcal{S}^n) \geq R$. Since (3) implies that $\Pr[\mathcal{A}] > 1-\epsilon$, we notice that R is a valid candidate for $\bar{R}$ in (5). Hence, $C_1 \leq C_2$.

To show that $C_2 \leq C_1$, it is equivalent to show that for any $\bar{R}$ in (5) that satisfies $\Pr[\mathcal{S}^n : C(\mathcal{S}^n) > \bar{R} - \epsilon] > 1 - \epsilon$, there exists a sequence of $\left(2^{n\bar{R}}, n\right)$ codebooks, as given by Definition 1, that satisfies (3). Then, by taking the supremum over $\bar{R}$, we get $C_2 \leq C_1$. Fix an $\epsilon > 0$. For any $\bar{R}$ which satisfies

$$\Pr[\mathcal{S}^n : C(\mathcal{S}^n) > \bar{R} - \epsilon] > 1 - \epsilon \tag{37}$$

define the set of $\mathcal{S}^n$ that satisfies (37) for this $\bar{R}$ as $\mathcal{B}$. For any $\mathcal{S}^n \in \mathcal{B}$, $C(\mathcal{S}^n) > \bar{R} - \epsilon$, which means that there exists a sequence of codebooks $\mathcal{CK}(\mathcal{S}^n)$, for this particular $\mathcal{S}^n$, such that for sufficiently large n, $P_e(\mathcal{S}^n, f^n(\mathcal{S}^n), g^n(\mathcal{S}^n))$ as defined in (4) with $R(\mathcal{S}^n) = \bar{R}$, is less than ϵ. Hence, we form a sequence of $\left(2^{n\bar{R}}, n\right)$ codebooks in Definiton 1 as follows: for $\mathcal{S}^n \in \mathcal{B}$, let the codebook be $\mathcal{CK}(\mathcal{S}^n)$, and for $\mathcal{S}^n \notin \mathcal{B}$, let the sequence of codebooks be arbitrary with rate $\bar{R}$. For this sequence of $\left(2^{n\bar{R}}, n\right)$ codebooks in Definition 1, since (37) implies that $\Pr[\mathcal{B}] > 1-\epsilon$, (3) is satisfied, i.e., $\bar{R}$ is achievable in Definition 1. Hence, we have $C_2 \leq C_1$. Therefore, the two different capacity definitions yield the same number C, i.e., $C_1 = C_2 \triangleq C$.

Appendix B
Proof of Lemma 5

From the independence of S_i and (W, S^{i-1}, Y^{i-1}), and the Markov chain $(W, S^{i-1}, Y^{i-1}) - S_i, X_i - Y_i$, it follows that

$$(W, S^{i-1}, Y^{i-1}) - X_i - Y_i. \tag{38}$$

Fano's inequality implies that for reliable communications we require

$$\begin{aligned}
nR &\leq I(W; Y^n) \\
&\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|W, Y^{i-1}) \\
&\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|W, Y^{i-1}, S^{i-1}) \\
&=^{(a)} \sum_{i=1}^{n} H(Y_i) - H(Y_i|Y^{i-1}, W, S^{i-1}, X_i) \\
&=^{(b)} \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \\
&\leq \max_i I(X_i; Y_i)
\end{aligned} \tag{39}$$

where (a) follows by (6) and (b) follows by (38).

References

[1] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problemy Peredachi Informatsii*, vol. 9, no. 1, pp. 19–31, 1980.

[3] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai(Shitz), "On the capacity of interference channels with a cognitive transmitter," *European Trans. on Telecommunications, invited*, vol. 19, pp. 405–420, Apr. 2008.

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

[5] S. Sridharan and S. Vishwanath, "On the capacity of a class of MIMO cognitive radios," in *IEEE Information Theory Workshop (ITW 2007), Lake Tahoe, CA*, Sep. 2007.

[6] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Trans. Inf. Theory, submitted, http://www.arxiv.org/pdf/cs.IT/0604107.pdf*, 2006.

[7] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731 – 739, Sep. 1983.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[9] A. D. Wyner, "Bounds on the rate-distortion function for stationary sources with memory," *IEEE Trans. Inf. Theory*, vol. 17, no. 5, pp. 508 – 513, Sep. 1971.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley Sons, Inc., 1991.

[11] C. E. Shannon, "Channels with side information at the transmitter," in *IBM Research and Development*, vol. 2, 1958, pp. 289–293.

[12] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[13] R. Tandra and A. Sahai, "Is interference like noise when you know its codebook?" in *IEEE International Symposium on information Theory*, Seattle, WA, July 2006.