Multi-user Information Theory October 23rd, 2013

Lecture 1

Lecturer: Haim Permuter *Scribe: Oron Sabag*

I. Introduction

This lecture gives an overall review and motivation for the Interference Channel(IC). The interference channel is based on a network consisting in its general form N senders and N receivers. Specifically, in our class we study the case where $N = 2$. There exists a one-to-one correspondence between senders and receivers. Each sender only wants to communicate with its corresponding receiver, and each receiver only cares about the information form its corresponding sender. However, each channel interferes the others.

Motivation for this model can be found in satellite communication. For instance, two satellites send information to its corresponding ground station simultaneously. Each ground station can receive the signals from both of the two satellites and its communication is interfered by the other pair's communication. The interference channel models also a wireless communication and a wired communication on a **twisted pair due to e** .

The IC was first studied in 1974 by Ahlswede in [1], where inner and outer bounds were derived. Later, Han and Kobayashi derived in [2] the best known-inner bound on the capacity region of the DM-IC. This inner bound was found to be tight for any any special case which has a capacity region; such that deterministic IC and strong interference IC. However, this channel has not been solved in general case even in the general Gaussian case.

In this lecture note, Section II describes the problem definition of the IC, and Section III includes 3 capacity regions regarding the *Cognitive IC* including detailed proofs.

II. Problem definition

The IC is described in Fig. 1. The DM-IC model $(\mathcal{X}_1, \mathcal{X}_2, P_{Y_1,Y_2|X_1,X_2}, \mathcal{Y}_1, \mathcal{Y}_2)$, consists of four finite alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2$ and a collection of conditional pmfs $P_{Y_1,Y_2|X_1,X_2}$

on $\mathcal{Y}_1, \mathcal{Y}_2$.

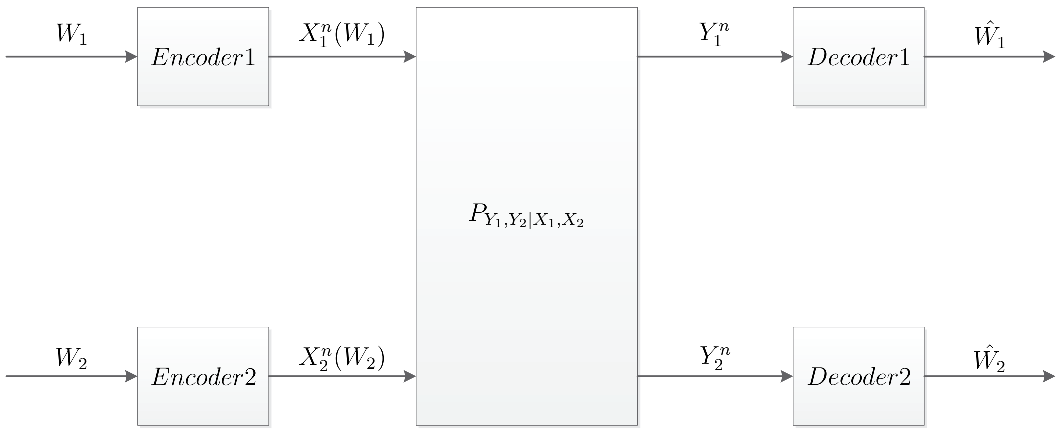

Fig. 1. Interference Channel for two users.

Definition 1 (Code for the IC) A $(2^{nR_1}, 2^{nR_2}, n)$ *code* for the interference channel consists of:

- Two message sets $\mathcal{W}_1 = \{1, \dots, 2^{nR_1}\}$ and $\mathcal{W}_2 = \{1, \dots, 2^{nR_2}\}$
- Encoding function $g_j : \mathcal{W}_j \to \mathcal{X}_j^n$, for $j = 1, 2$
- Decoding function $\phi_j : \mathcal{Y}_j^n \to \hat{\mathcal{W}}_j$, for $j = 1, 2$

We assume that the message pair (W_1, W_2) is uniformly distributed over $\{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}$. The *average probability of error* is defined as $P_e^{(n)} = \Pr((W_1, W_2) \neq (\hat{W}_1, \hat{W}_2))$. A rate pair (R_1, R_2) is said to be *achievable* for the DM-IC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$.

The *capacity region* is defined as the closure of the set of achievable rate pairs (R_1, R_2).

III. Cognitive Interference Channel

A. Deterministic Cognitive Interference Channel

In this section we study a special case for the IC, the *Deterministic-Cognitive IC*. The setup is described in Fig. 2. In this model we assume that the message W_2 is known at both encoders. Moreover, the channel is memoryless and induced by two deterministic function f_1 and f_2 with the input arguments (X_1, X_2), that is, $P_{Y_1,Y_2|X_1,X_2} = \mathbb{1}_{Y_1=f_1(X_1,X_2)} \cdot \mathbb{1}_{Y_2=f_2(X_1,X_2)}$.

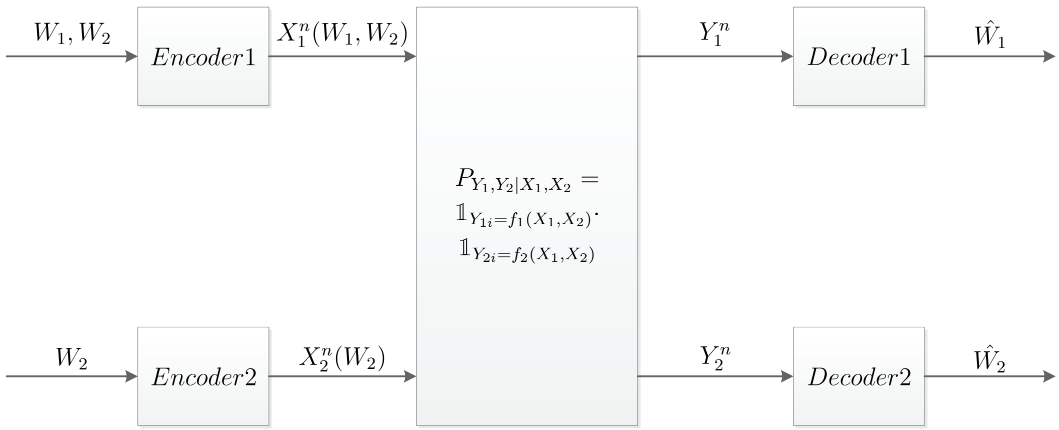

Fig. 2. Deterministic Cognitive Interference Channel for two users.

Theorem 1 (Capacity Region for the Deterministic-Cognitive IC) The capacity region is the set of rate pairs (R_1, R_2), such that:

$$\begin{aligned} R_1 &\leq H(Y_1|X_2), \\ R_2 &\leq H(Y_2|X_2) + I(X_2; Y_2), \\ R_1 + R_2 &\leq H(Y_1, Y_2|X_2) + I(X_2; Y_2), \end{aligned} \tag{1}$$

for some joint distribution $P_{X_1,X_2} \mathbb{1}_{Y_1=f_1(X_1,X_2)} \mathbb{1}_{Y_2=f_2(X_1,X_2)}$.

Proof:

Achievability: The achievability comprises of two steps; first, we use the the deterministic broadcast channel (BC) coding scheme to transmit from encoder 1 at rates (R_1, R'_2), then encoder 2 transmits additional information at rate R''_2. Combining both steps, we conclude that the rate pair $(R_1, R'_2 + R''_2)$ is achievable.

Let us remind to the reader the capacity region of the deterministic BC where non-causal side information S^n is given to the encoder:

$$R_1 \leq H(Y_1|S),$$

$$R_2 \leq H(Y_2|S),$$

$$R_1 + R_2 \leq H(Y_1, Y_2|S).$$

Encoder 1 is using the coding scheme of the deterministic BC at rates of rates (R_1, R_2'), but where the considered SI is X_2^n. Then, we use a simple point to point result for the rate can be achieved from encoder 2 to decoder 2,i.e. $R_2'' \leq I(X_2; Y_2)$. Combining these two coding schemes, we achieve the region:

$$R_1 \leq H(Y_1|X_2), \tag{2}$$

$$R_2 \leq H(Y_2|X_2) + I(X_2; Y_2), \tag{3}$$

$$R_1 + R_2 \leq H(Y_1, Y_2|X_2) + I(X_2; Y_2). \tag{4}$$

Converse: For the converse part, we assume that there exists a code $(2^{nR_1}, 2^{nR_2}, n)$ such that $\lim_{n\to\infty} P_e^{(n)} = 0$.

For the rate R_1, consider

$$\begin{aligned}
nR_1 &= H(W_1) \\
&\overset{(a)}{=} H(W_1|W_2) \\
&\overset{(b)}{=} H(W_1|W_2, X_2^n) \\
&\overset{(c)}{=} H(W_1, Y_1^n|W_2, X_2^n) \\
&= H(Y_1^n|W_2, X_2^n) + H(W_1|Y_1^n, W_2, X_2^n) \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} H(Y_{1i}|X_{2i}) + n\epsilon_n
\end{aligned}$$

where:

(a) follows from the fact that the messages W_1 and W_2 are independent;

(b) follows from the fact that X_2^n is a deterministic function of W_2;

(c) follows from the deterministic channel characterization;

(d) follows from Fano's inequality, i.e. $H(W_1|Y_1^n, W_2, X_2^n) \leq n\epsilon_n$ and the fact that conditioning reduces entropy.

For the rate R_2, consider

$$\begin{aligned}
nR_2 &= H(W_2) \\
&\overset{(a)}{\leq} I(W_2; Y_2^n) + n\epsilon_n \\
&\overset{(b)}{\leq} H(Y_2^n) + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} H(Y_{2i}) + n\epsilon_n
\end{aligned}$$

where:

(a) follows from Fano's inequality, i.e. $H(W_2|Y_2^n) \leq n\epsilon_n$;

(b) follows from the non-negativity of the term $H(Y_2^n|W_2)$;

(c) follows from the fact that conditioning reduces entropy.

For the sum rate $R_1 + R_2$, consider

$$\begin{aligned}
n(R_1 + R_2) &= H(W_1, W_2) \\
&= H(W_1|W_2) + H(W_2) \\
&\overset{(a)}{\leq} H(W_1|W_2, X_2^n) + I(W_2; Y_2^n) + n\epsilon_n \\
&\overset{(b)}{=} H(Y_1^n, W_1|W_2, X_2^n) + H(Y_2^n) - H(Y_2^n|W_2) + n\epsilon_n \\
&\overset{(c)}{\leq} H(Y_1^n, W_1|W_2, X_2^n) + H(Y_2^n) + n\epsilon_n \\
&\overset{(d)}{\leq} H(Y_1^n|W_1, W_2, X_2^n) + H(Y_2^n) + 2n\epsilon_n \\
&\overset{(e)}{=} H(Y_1^n|W_1, W_2, X_2^n, Y_2^n) + H(Y_2^n) + 2n\epsilon_n \\
&\overset{(f)}{\leq} \sum_{i=1}^{n} H(Y_{1i}|X_{2i}, Y_{2i}) + H(Y_{2i}) + 2n\epsilon_n
\end{aligned}$$

where:

(a) follows from the fact that X_2^n is a deterministic function of W_2 and Fano's inequality, i.e. $H(W_2|Y_2^n) \leq n\epsilon_n$;

(b) follows from the deterministic channel characterization;

(c) follows from the non-negativity of the term $H(Y_2^n|W_2)$;

(d) follows from Fano's inequality, i.e. $H(W_1|Y_1^n, W_2, X_2^n) \leq n\epsilon_n$;

(e) follows from the fact that Y_2^n is a deterministic function of (W_1, W_2);

(f) follows from the fact that conditioning reduces entropy.

■

B. Semi-Deterministic Cognitive IC with state known at the cognitive user

The setup is described in Fig. 3, the *Semi-Deterministic* (SD) channel is characterized by the channel distribution $P_{Y_1,Y_2|X_1,X_2,S} = \mathbf{1}_{y_1=f(x_1,x_2,s)}P(y_2|x_1,x_2,s)$ for each time instance. Two independent messages M_1, M_2 are distributed uniformly in the set $\{1,\ldots,2^{nR_1}\} \times \{1,\ldots,2^{nR_2}\}$ and the state of the channel is $i.i.d.$ and distributed according to $P(s)$, independently from each messages. Encoder 1 transmits the signal X_1^n to the channel, based on both messages and the non-causal side information sequence S^n. Encoder 2 has access to the message M_2 only, and transmits the signal X_2^n to the channel. Based on the output Y_i^n, Decoder i decodes the message M_i, where $i \in \{1,2\}$.

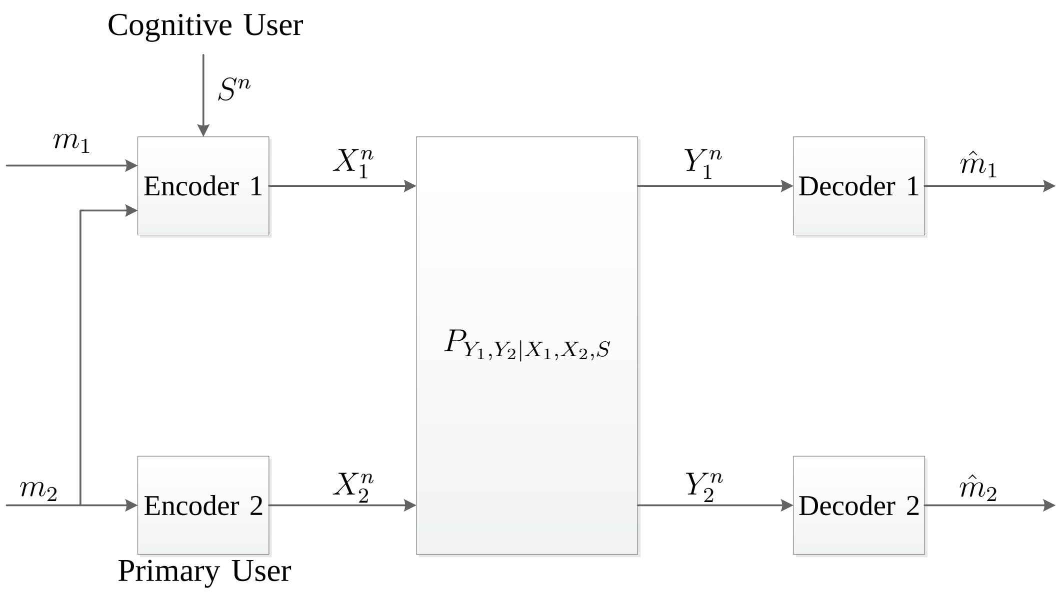

Fig. 3. Cognitive interference channel, where the cognitive transmitter knows the state non causally.

Definition 2 A $(2^{nR_1}, 2^{nR_2}, n)$ *code* of blocklength n for the setting in Fig. 3 consisting of two encoding functions

$$f_{e1} : \{1,...,2^{nR_1}\} \times \{1,...,2^{nR_2}\} \times \mathcal{S}^n \mapsto \mathcal{X}_1^n,$$

$$f_{e2} : \{1, ..., 2^{nR_2}\} \mapsto \mathcal{X}_2^n, \tag{5}$$

and two decoding functions

$$g_{d1} : \mathcal{Y}_1^n \mapsto \{1, ..., 2^{nR_1}\},$$
$$g_{d2} : \mathcal{Y}_1^n \mapsto \{1, ..., 2^{nR_2}\}. \tag{6}$$

Let us denote by $\hat{M}_1(Y_1^n)$ and $\hat{M}_2(Y_2^n)$ the outputs of Decoder 1 and 2, respectively.

Definition 3 The probability of error, $P_e^{(n)}$ of a code of blocklength n is defined as

$$P_e^{(n)} = \Pr\{M_1 \neq \hat{M}_1(Y_1^n) \text{ or } M_2 \neq \hat{M}_2(Y_1^n)\}. \tag{7}$$

We use the standard definition of an achievable pair-rate and the capacity region.

The next Theorem was introduced and proved in [4].

Theorem 2 (Capacity of the SD Cognitive IC with State known to the Cognitive User [4]) The capacity region of the semi-deterministic interference channel with state known to the cognitive user is the set of all pair rates (R_1, R_2) that satisfies

$$\begin{aligned} R_1 &\leq H(Y_1|S, X_2) \\ R_2 &\leq I(U, X_2; Y_2) - I(U, X_2; S) \\ R_1 + R_2 &\leq H(Y_1|S, X_2, U) + I(U, X_2; Y_2) - I(U, X_2; S) \end{aligned} \tag{8}$$

for some joint distribution of the form $P(s)P(x_2)P(x_1, u|x_2, s)\mathbf{1}_{y_1=f(x_1,x_2,s)}P(y_2|x_1, x_2, s)$.

Theorem 3 (Capacity of the Deterministic Cognitive IC with State known to the Cognitive User) The capacity region of the deterministic interference channel with state known to the cognitive user is the set of all pair rates (R_1, R_2) that satisfies

$$\begin{aligned} R_1 &\leq H(Y_1|S, X_2) \\ R_2 &\leq H(Y_2) - I(Y_2, X_2; S) \\ R_1 + R_2 &\leq H(Y_1|S, X_2, Y_2) + H(Y_2) - I(Y_2, X_2; S) \end{aligned} \tag{9}$$

for some joint distribution of the form $P(s)P(x_2)P(x_1|x_2, s)\mathbf{1}_{y_1=f_1(x_1,x_2,s)}\mathbf{1}_{y_2=f_2(x_1,x_2,s)}$.

The achievability of Theorem 3 follows in a straightforward manner from Theorem 2 by replacing U with Y_2. It is also possible to write (9) as

$$\begin{aligned} R_1 &\leq H(Y_1|S,X_2) \\ R_2 &\leq H(Y_2|S,X_2) + I(X_2;Y_2) \\ R_1 + R_2 &\leq H(Y_1|S,X_2) + H(Y_2|S,X_2,Y_1) + I(Y_2;X_2). \end{aligned} \tag{10}$$

Note that the capacity region defined only by the entropy expressions of (10) are the capacity region of the deterministic BC with state (X_2, S) known non-causally at the encoder, and the additional rate $I(Y_2;X_2)$ is the point-to-point capacity of the primary user where (S, X_2, Y_1) are treated as noise.

Proof of Theorem 2:

Sketch of achievability: The main idea in the achievability is to split message M_2 into two parts with rates R_2' and R_2''. Then send the bit-rate R_2'' via a point-to-point channel to Decoder 2, where X_1 and S are treated as noise. The bit-rates R_1 and R_2' are sent from Encoder 1 using a semi-deterministic broadcast channel coding scheme with known state and the encoder [3], where the state is (S, X_2).

Here is a more detailed description. Fix a joint distribution $P(s)P(x_2)P(x_1,u|x_2,s)$. Split the message M_2 into two messages M_2' and M_2'' with rates R_2' and R_2'', respectively, such that

$$R_2 = R_2' + R_2''. \tag{11}$$

Code design: Generate $2^{nR_2''}$ random codewords X_2^n using i.i.d. $p(x_2)$. Generate a random code $(2^{nR_1}, 2^{nR_2'}, n)$ for a semi deterministic BC with state known at the encoder non causally as described in [3] where the state is (S^n, X_2^n). The outputs of the BC are Y_1 and (Y_2, X_2).

Encoding: Map the message M_2'' to a codeword X_2^n and transmit it. Map the message pair (M_1, M_2') to X_1^n where the state is $(S^n, X_2^n(M_2''))$, and transmit X_1^n.

Decoding: Decoder 2 receives Y_2^n and uses point-to-point decoding in order to decode $\hat{M}_2''$. Then Decoder 1 and Decoder 2 uses semideterministic BC with state known at the encoder to decoder $\hat{M}_1$ and $\hat{M}_2'$ at decoder 1 and 2, respectively. The state of the semideterministic BC is (X_2^n, S^n) and Decoder 1 uses Y_1^n to decode $\hat{M}_1$ and Decoder 2 uses $(Y_2^n, X_2(\hat{M}_2''))$ to decode $\hat{M}_2'$.

Error analysis: We would like to show that if the rate-pair (R_1, R_2) satisfies (8) with a strict inequality then as n goes to infinity $P_e^{(n)}$ goes to zero.

First note that if

$$R_2'' < I(X_2; Y_2), \tag{12}$$

then Decoder 2 would be able to decode M_2'' with a probability of error that goes to zero.

Now considering a semi deterministic BC with state (the state is $(X_2^n(M_2''), S^n)$) known at the encoder, where the first decoder obtain Y_1^n and the second decoder obtains $(Y_2^n, X_2^n(\hat{M}_2'')$. Using the achievability from [3] if (R_1, R_2') satisfies

$$\begin{aligned} R_1 &< H(Y_1|S, X_2) \\ R_2' &< I(U; Y_2, X_2) - I(U; S, X_2) \\ R_1 + R_2' &< H(Y_1|S, X_2) + I(U; Y_2, X_2) - I(U; S, X_2, Y_1) \end{aligned} \tag{13}$$

then Decoder 1 and 2 would be able to decode M_1 and M_2', respectively, with a probability of error that goes to zero. Using Fourier-Mozkin elimination on (11), (12) and (14) we obtain

$$\begin{aligned} R_1 &< H(Y_1|S, X_2) \\ R_2 &< I(U; Y_2, X_2) - I(U; S, X_2) + I(X_2; Y_2) \\ R_1 + R_2 &< H(Y_1|S, X_2) + I(U; Y_2, X_2) - I(U; S, X_2, Y_1) + I(X_2; Y_2). \end{aligned} \tag{14}$$

Finally, using simple chain rules and the fact that X_2 is independent of S we obtain

$$\begin{aligned} & I(U; Y_2, X_2) - I(U; S, X_2) + I(X_2; Y_2) \\ &= I(U; Y_2|X_2) - I(U; S|X_2) + I(X_2; Y_2) \\ &= I(U, X_2; Y_2) - I(U; S|X_2) \\ &= I(U, X_2; Y_2) - I(U, X_2; S), \end{aligned} \tag{15}$$

and

$$H(Y_1|S, X_2) + I(U; Y_2, X_2) - I(U; S, X_2, Y_1) + I(X_2; Y_2)$$

$$
\begin{aligned}
&= H(Y_1|S,X_2) + I(U;Y_2|X_2) - I(U;S,Y_1|X_2) + I(X_2;Y_2) \\
&= H(Y_1|S,X_2) + I(X_2,U;Y_2) - I(U;S,Y_1|X_2) \\
&= H(Y_1|S,X_2) + I(X_2,U;Y_2) - I(U;S|X_2) - I(U;Y_1|X_2,S) \\
&= H(Y_1|S,X_2,U) + I(X_2,U;Y_2) - I(U,X_2;S), \qquad (16)
\end{aligned}
$$

and this prove that (14) is identical to (8). ■

Proof of Converse: Let us fix a code $(2^{nR_1}, 2^{nR_2}, n)$ with a probability of error, $P_e^{(n)}$. Now consider the following inequalities,

$$
\begin{aligned}
nR_1 &= H(M_1) \\
&= H(M_1|S^n, M_2) \\
&\overset{(a)}{\leq} I(M_1;Y^n|S^n,M_2) + n\epsilon_n, \\
&\leq H(Y^n|S^n,M_2) + n\epsilon_n, \\
&= H(Y_i|Y^{i-1},S^n,M_2,X_{2,i}(M_2)) + n\epsilon_n, \\
&\leq \sum_{i=1}^{n} H(Y_i|S_i,X_{2,i}) + n\epsilon_n, \qquad (17)
\end{aligned}
$$

where step (a) follows from Fano's inequality and $\epsilon_n \triangleq (R_1+R_2)P_e^{(n)} + \frac{1}{n}$. The second set of inequalities is very similar to the converse of point-to-point channel with non causal state known at the encoder (Gelfand-Pinsker).

$$
\begin{aligned}
nR_2 =& H(M_2) \\
\overset{(a)}{\leq}& I(M_2;Y_2^n) - I(M_2;S^n) + n\epsilon_n \\
=& \sum_{i=1}^{n} I(M_2;Y_{2,i}|Y_2^{i-1}) - I(M_2;S_i|S_{i+1}^n) + n\epsilon_n, \\
=& \sum_{i=1}^{n} I(M_2,S_{i+1}^n;Y_{2,i}|Y_2^{i-1}) - I(S_{i+1}^n;Y_{2,i}|Y_2^{i-1},M_2) \\
& - I(M_2,Y_2^{i-1};S_i|S_{i+1}^n) + I(Y_2^{i-1};S_i|S_{i+1}^n,M_2) + n\epsilon_n \\
\overset{(b)}{=}& \sum_{i=1}^{n} I(M_2,S_{i+1}^n;Y_{2,i}|Y_2^{i-1}) - I(M_2,Y_2^{i-1},S_{i+1}^n;S_i) + n\epsilon_n
\end{aligned}
$$

$$\leq \sum_{i=1}^{n} I(M_2, S_{i+1}^n, Y_2^{i-1}; Y_{2,i}) - I(M_2, Y_2^{i-1}, S_{i+1}^n; S_i) + n\epsilon_n,$$

$$\stackrel{(c)}{=} \sum_{i=1}^{n} I(X_{2,i}, M_2, S_{i+1}^n, Y_2^{i-1}; Y_{2,i}) - I(X_{2,i}, M_2, Y_2^{i-1}, S_{i+1}^n; S_i) + n\epsilon_n,$$

$$\stackrel{(d)}{=} \sum_{i=1}^{n} I(X_{2,i}, V_i; Y_{2,i}) - I(X_{2,i}, V_i; S_i) + n\epsilon_n, \tag{18}$$

where step (a) follows from Fano's inequality, step (b) from Csiszar sum identity, i.e., $\sum_{i=1}^n I(A^{i-1}; B_i|B_{i+1}^n) = \sum_{i=1}^n I(B_{i+1}^n; A_i|A^{i-1})$, step (c) from the fact that $X_{2,i}$ is a function of M_2 and step (d) from defining

$$V_i \triangleq (M_2, Y_2^{i-1}, S_{i+1}^n). \tag{19}$$

To prove the converse of the third inequality in (8) we would use the following identity which follows simply from the chain rule of mutual information.

$$H(Y_1|S, X_2, U) + I(U, X_2; Y_2) - I(U, X_2; S) \tag{20}$$

$$= H(Y_1|S, X_2, U) - H(Y_1|S) + H(Y_1|S) + I(U, X_2; Y_2) - I(U, X_2; S)$$

$$= -I(Y_1; X_2, U|S) + H(Y_1|S) + I(U, X_2; Y_2) - I(U, X_2; S)$$

$$= H(Y_1|S) + I(U, X_2; Y_2) - I(U, X_2; S, Y_1) \tag{21}$$

Now consider,

$$n(R_1 + R_2) \tag{22}$$

$$= H(M_1) + H(M_2)$$

$$\stackrel{(a)}{\leq} H(M_1|S^n) + I(M_2; Y_2^n) - I(M_2; M_1, S^n) + n\epsilon_n$$

$$= H(M_1|S^n) + I(M_2; Y_2^n) - I(M_2; M_1, S^n, Y_1^n) + I(M_2; Y_1^n|M_1, S^n) + n\epsilon_n$$

$$\stackrel{(b)}{=} H(M_1, Y_1^n|S^n) + I(M_2; Y_2^n) - I(M_2; M_1, S^n, Y_1^n) + n\epsilon_n$$

$$\stackrel{(c)}{\leq} H(Y_1^n|S^n) + I(M_2; Y_2^n) - I(M_2; S^n, Y_1^n) + 2n\epsilon_n$$

$$= H(Y_1^n, S^n) - H(S^n) + I(M_2; Y_2^n) - I(M_2; S^n, Y_1^n) + 2n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_{1,i}, S_i|Y_{1,i+1}^n, S_{i+1}^n) - H(S_i) + I(M_2; Y_{2,i}|Y_2^{i-1}) - I(M_2; Y_{1,i}, S_i|Y_{1,i+1}^n, S_{i+1}^n) + 2n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_{1,i}, S_i) - H(Y_{1,i}, S_i) + H(Y_{1,i}, S_i | Y_{1,i+1}^n, S_{i+1}^n) - H(S_i)$$

$$+ I(M_2; Y_{2,i} | Y_2^{i-1}) - I(M_2; Y_{1,i}, S_i | Y_{1,i+1}^n, S_{i+1}^n) + 2n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_{1,i}|S_i) - I(Y_{1,i}, S_i; Y_{1,i+1}^n, S_{i+1}^n) + I(M_2; Y_{2,i}|Y_2^{i-1}) - I(M_2; Y_{1,i}, S_i | Y_{1,i+1}^n, S_{i+1}^n) + 2n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_{1,i}|S_i) + I(M_2; Y_{2,i}|Y_2^{i-1}) - I(M_2, Y_{1,i+1}^n, S_{i+1}^n; Y_{1,i}, S_i) + 2n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_{1,i}|S_i) + I(M_2, Y_{1,i+1}^n, S_{i+1}^n; Y_{2,i}|Y_2^{i-1}) - I(Y_{1,i+1}^n, S_{i+1}^n; Y_{2,i}|M_2, Y_2^{i-1})$$

$$- I(M_2, Y_{1,i+1}^n, S_{i+1}^n, Y_2^{i-1}; Y_{1,i}, S_i) + I(Y_2^{i-1}; Y_{1,i}, S_i | M_2, Y_{1,i+1}^n, S_{i+1}^n) + 2n\epsilon_n$$

$$\overset{(d)}{=} \sum_{i=1}^{n} H(Y_{1,i}|S_i) + I(M_2, Y_{1,i+1}^n, S_{i+1}^n; Y_{2,i}|Y_2^{i-1}) - I(M_2, Y_{1,i+1}^n, S_{i+1}^n, Y_2^{i-1}; Y_{1,i}, S_i) + 2n\epsilon_n$$

$$\leq \sum_{i=1}^{n} H(Y_{1,i}|S_i) + I(X_{2,i}, M_2, Y_{1,i+1}^n, S_{i+1}^n, Y_2^{i-1}; Y_{2,i}) - I(X_{2,i}, M_2, Y_{1,i+1}^n, S_{i+1}^n, Y_2^{i-1}; Y_{1,i}, S_i) + 2n\epsilon_n$$

$$\overset{(e)}{=} \sum_{i=1}^{n} H(Y_{1,i}|S_i) + I(X_{2,i}, V_i, T_i; Y_{2,i}) - I(X_{2,i}, V_i, T_i; Y_{1,i}, S_i) + 2n\epsilon_n, \tag{23}$$

where (a) and (c) follows from Fano's inequality and the independence of M_1, M_2 and S^n, from the fact that Y_1^n is a deterministic function of M_1, M_2, S^n), (d) from Csiszar sum identity, and (e) from the definition of V_i which is given in (19) and the definition of T_i,

$$T_i \triangleq Y_2^{i-1}. \tag{24}$$

Now we are using the trick that was introduced in [3] to overcome the fact that the auxiliary T is not present in the converse of the second inequality given in (18). We need to find a U for which

$$I(X_2, V; Y_2) - I(X_2, V; S) \leq I(X_2, U; Y_2) - I(X_2, U; S)$$

$$H(Y_1|S) + I(X_2, V, T; Y_2) - I(X_2, V, T; Y_1, S) \leq H(Y_1|S) + I(X_2, U; Y_2) - I(X_2, U; Y_1, S). \tag{25}$$

As in [3] we will show there always exists such a U. Note that if

$$I(X_2, T; Y_2|V) - I(X_2, T; Y_1, S|V) \leq 0, \tag{26}$$

then we can choose $U = V$, and (25) will hold, and if

$$I(X_2, T; Y_2|V) - I(X_2, T; S|V) \geq 0 \tag{27}$$

then we can $U = (V, T)$, and (25) will hold. Furthermore, note that one of the conditions (26) or (27) will always hold, therefore there exists a choice of U for which (25) holds.

■

Converse proof of Theorem 3: Let us fix a code $(2^{nR_1}, 2^{nR_2}, n)$ with a probability of error, $P_e^{(n)}$. The inequality

$$nR_1 \leq \sum_{i=1}^{n} H(Y_i|S_i, X_{2,i}) + n\epsilon_n, \tag{28}$$

follows from identical steps as (17). Now consider the rate R_2,

$$\begin{aligned}
nR_2 &= H(M_2) \\
&= H(M_2|S^n) \\
&\leq H(M_2, Y_2^n|S^n) - H(M_2, Y_2^n) + H(M_2, Y_2^n) \\
&\overset{(a)}{=} H(Y_2^n) - I(M_2, Y_2^n; S^n) + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} H(Y_{2,i}) - H(S_i) + H(S_i|X_{2,i}, Y_{2,i}) + n\epsilon_n, \\
&\leq \sum_{i=1}^{n} H(Y_{2,i}) - I(S_i; X_{2,i}, Y_{2,i}) + n\epsilon_n,
\end{aligned} \tag{29}$$

where (a) follows from Fano's Inequality and defining $\epsilon_n \triangleq (R_1 + R_2)P_e^{(n)} + \frac{1}{n}$, and (b) from the facts that S^n is distributed i.i.d., conditioning reduces entropy and $X_{2,i}$ is a function of M_2. For the sum rate consider,

$$\begin{aligned}
n(R_1 + R_2) &= H(M_1, M_2) \\
&= H(M_1, M_2, S^n) - H(S^n) \\
&\overset{(a)}{=} H(Y_1^n, Y_2^n, M_1, M_2, S^n) - H(S^n), \\
&\overset{(b)}{\leq} H(Y_1^n, Y_2^n, S^n) - H(S^n) + n\epsilon_n, \\
&\leq H(Y_2^n) + H(Y_1^n, S^n|Y_2^n) - H(S^n) + n\epsilon_n,
\end{aligned}$$

$$
\begin{aligned}
&\leq H(Y_2^n) + H(X_2^n, Y_1^n, S^n | Y_2^n) - H(S^n) + n\epsilon_n, \\
&\overset{(c)}{\leq} H(Y_2^n) + H(Y_1^n, S^n | Y_2^n, X_2^n) - H(S^n) + 2n\epsilon_n, \\
&\leq \sum_{i=1}^{n} H(Y_{2,i}) + H(Y_{1,i}, S_i | Y_{2,i}, X_{2,i}) - H(S_i) + 2n\epsilon_n, \\
&= \sum_{i=1}^{n} H(Y_{2,i}) + H(Y_{1,i} | S_i, X_{2,i}, Y_{2,i}) - I(Y_{2,i}, X_{2,i}; S_i) + 2n\epsilon_n, \qquad (30)
\end{aligned}
$$

where:

(a) follows from the deterministic channel characterization;

(b) follows from Fano's inequality, i.e. $H(M_1, M_2 | Y_1^n, Y_2^n, S^n) \leq n\epsilon_n$;

(c) follows from Fano's inequality, i.e. $H(X_2^n | Y_2^n) \leq n\epsilon_n$.

■

REFERENCES

[1] R. Ahlswede, The Capacity Region of a Channel with Two Senders and Two Receivers, The Annals of Probability, Vol. 2, No. 5, pp. 805-814, Oct. 1974.

[2] T. S. Han and K. Kobayashi, A New Achievable Rate Region for the Interference Channel, IEEE Trans. Information theory, Vol. IT-27, No. 1, pp. 49-60, Jan. 1981.

[3] A. Lapidoth,W. Ligong, The State-Dependent Semideterministic Broadcast Channel, Information Theory, IEEE Trans. Information theory, vol.59, no.4, pp.2242-2251, April 2013.

[4] M. Kazemiand A. Vosoughi, On the capacity of the state-dependent cognitive interference channel, In Proc. IEEE International Symposium on Information Theory (ISIT), July 2013.