Binary Fading Interference Channel with No CSIT
Alireza Vahid, Mohammad Ali Maddah-Ali, and A. Salman Avestimehr

Abstract

We characterize the capacity region of the two-user Binary Fading Interference Channel where the transmitters have no knowledge of the channel state information. We show that the entire capacity region is achieved by applying point-to-point erasure codes with appropriate rates at each transmitter, and using either treat-interference-as-erasure or interference-decoding at each receiver, based on the channel parameters. The result is obtained by developing a novel outer-bound that has three main steps. We first create a contracted channel that has fewer states compared to the original channel, in order to make the analysis tractable. Using a Correlation Lemma, we then show that an outer-bound on the capacity region of the contracted channel also serves as an outer-bound for the original channel. Finally, using a Conditional Entropy Leakage Lemma, we derive our outer-bound on the capacity region of the contracted channel, and show that it coincides with the achievable region by either treat-interference-as-erasure or interference-decoding at each receiver. We also show that having access to delayed local knowledge of the channel state information, does not enlarge the capacity region.

Index Terms

Interference channel, binary fading, capacity, channel state information, no CSIT, delayed local CSIT, packet collision.

I. INTRODUCTION

The two-user Interference Channel (IC) introduced in [2], is a canonical example to study the impact of interference in communication networks. There exists an extensive body of work on this problem (e.g., [3]–[5]). In this work, we focus on a specific configuration of this network, named the two-user Binary Fading Interference Channel (BFIC) as depicted in Fig. 1 in which the channel gains at each time instant are either 0 or 1 according to some distribution. The input-output relation of this channel at time instant $t$ is given by

$$Y_i[t] = G_{ii}[t]X_i[t] \oplus G_{\bar{i}i}[t]X_{\bar{i}}[t], \quad i = 1, 2,$$

where $\bar{i} = 3 - i$, $G_{ii}[t], G_{\bar{i}i}[t] \in \{0, 1\}$, and all algebraic operations are in $\mathbb{F}_2$.

Fig. 1. Two-user Binary Fading Interference Channel (BFIC).

The motivation for studying the Binary Fading Interference Channel is twofold. First, as demonstrated in [6], it provides a simple, yet useful physical layer abstraction for wireless packet networks in which whenever a collision occurs, the receiver can store its received analog signal and utilize it for decoding the packets in future (for example,
by successive interference cancellation techniques). In this context, the binary fading model is motivated by a shadow fading environment, in which each link is either “on” or “off” (according to the shadow fading distribution), and the multiple access (MAC) is modeled such that if two signals are transmitted simultaneously, and the links between the corresponding transmitters and the receiver are not in deep fade, then a linear combination of the signals is available to the receiver. The study of the BFIC in [6] has led to several coding opportunities that can be utilized by the transmitters to exploit the available signal at the receivers for interference management.

The second motivation for studying the BFIC is that it can be a first step towards understanding the capacity of the fading interference channels with no knowledge of the channel state information at the transmitters (CSIT). This model was used in [7], [8] to derive the capacity of the one-sided interference channel (also known as Z-Channel). Motivated by the deterministic approach [9], a layered erasure broadcast channel model was introduced in [10] to approximate the capacity of fading broadcast channels. One can view our Binary Fading model as the model introduced in [8], [10] with one layer.

In this work, we consider the two-user BFIC, and we consider two scenarios about the available channel state information at the transmitters (CSIT). In the no CSIT case, the transmitters are only aware of the distributions of the channel gains but not the actual realizations; and in the delayed local CSIT case, we assume that each transmitter learns the channel realization to its corresponding receiver with unit delay. We characterize the capacity region of this problem, and we show that under the two assumptions, the capacity region remains the same.

To derive the outer-bound, we incorporate two key lemmas. The first lemma, the Conditional Entropy Leakage Lemma, establishes how much information is leaked from each transmitter to the unintended receiver. The second lemma, the Correlation Lemma, shows that if the channel gains are correlated under a given set of conditions, the capacity region cannot be smaller than the case of independent channel gains. Using the Correlation Lemma, we create a contracted channel that has fewer states as opposed to the original channel and hence, the problem becomes tractable. Then, using the Conditional Entropy Leakage Lemma, we derive an outer-bound on the capacity region of the contracted channel which in turn, serves as an outer-bound for the original channel.

For the achievability, we show that the entire capacity region is achieved by applying point-to-point erasure codes with appropriate rates at each transmitter, and using either treat-interference-as-erasure or interference-decoding at each receiver, based on the channel parameters. Thus, message splitting (such as Han-Kobayashi scheme) is not required. We further prove that the achievable region under no CSIT assumption, matches the outer-bound on the capacity region with delayed local CSIT. Thus, the capacity region under the two models would be identical.

The rest of the paper is organized as follows. In Section II we formulate our problem. In Section III we present our main results. Section IV is dedicated to deriving the outer-bound for the no CSIT assumption. We describe our achievability strategy in Section V. In Section VI we show that delayed local knowledge of the CSI, does not enlarge the capacity region. Section VII concludes the paper and describes future directions.

II. PROBLEM SETTING

We consider the two-user Binary Fading Interference Channel (BFIC) as illustrated in Fig. 1. The channel gain from transmitter \( Tx_i \) to receiver \( Rx_j \) at time instant \( t \), is denoted by \( G_{ij}[t] \), \( i, j \in \{1, 2\} \). We assume that the channel gains are either 0 or 1 (i.e. \( G_{ij}[t] \in \{0, 1\} \)), and they are distributed as independent Bernoulli random variables (independent from each other and over time). Furthermore, we consider the symmetric setting where

\[
G_{ii}[t] \overset{d}{=} \mathcal{B}(p_d), \quad G_{ij}[t] \overset{d}{=} \mathcal{B}(p_c),
\]

(2)

for \( 0 \leq p_d, p_c \leq 1 \), \( i = 3 - j \), and \( i = 1, 2 \). We define \( q_d = 1 - p_d \) and \( q_c = 1 - p_c \).

At each time instant \( t \), the transmit signal at \( Tx_i \) is denoted by \( X_i[t] \in \{0, 1\} \), \( i = 1, 2 \), and the received signal at \( Rx_i \) is given by

\[
Y_i[t] = G_{ii}[t]X_i[t] \oplus G_{ij}[t]X_j[t], \quad i = 1, 2,
\]

(3)

where all algebraic operations are in \( \mathbb{F}_2 \). Due to the nature of the channel gains, a total of 16 channel realizations may occur at any given time instant as given in Table I.

The channel state information (CSI) at time instant \( t \) is denoted by the quadruple

\[
G[t] = (G_{11}[t], G_{12}[t], G_{21}[t], G_{22}[t]).
\]

(4)
For a natural number $k$, we set
\[ G^k = [G[1], G[2], \ldots, G[k]]^T. \] (5)

Finally, we set
\[ G^k_{ii}X_i^t \oplus G^k_{ii}X_i^t = [G^t_{ii}[1]X_i[1] \oplus G^t_{ii}[1]X_i[1], \ldots, G^t_{ii}[t]X_i[t] \oplus G^t_{ii}[t]X_i[t]]^T. \] (6)

In this paper, we assume that receiver $i$ has instantaneous knowledge of $G_{ii}[t]$ and $G_{ii}[t]$ (i.e. the incoming links to receiver $i$), $i = 1, 2$. Furthermore, we consider two models for the available channel state information at the transmitters:

1) No CSIT: In this model, we assume that transmitters only know the distribution from which the channel gains are drawn, but not the actual realizations of them;

2) Delayed Local CSIT: In this model, we assume that at time instant $t$, transmitter $i$ has access to $G^t_{ii}$ besides the distribution from which the channel gains are drawn, $i = 1, 2$.

Consider the scenario in which Tx$_i$ wishes to reliably communicate message $W_i \in \{1, 2, \ldots, 2^nR_i\}$ to Rx$_i$ during $n$ uses of the channel, $i = 1, 2$. We assume that the messages and the channel gains are mutually independent and the messages are chosen uniformly. For each transmitter Tx$_i$, let message $W_i$ be encoded as $X^n_i$ using the encoding function $f_i(.)$ that depends on the available CSIT. Receiver Rx$_i$ is only interested in decoding $W_i$, and it will decode the message using the decoding function $\hat{Y}_i = \varphi_i(Y^n_i)$ that depends on the available channel state information at that receiver. An error occurs when $\hat{W}_i \neq W_i$. The average probability of decoding error is given by
\[ \lambda_{i,n} = E[P[\hat{W}_i \neq W_i]], \quad i = 1, 2, \] (7)

and the expectation is taken with respect to the random choice of the transmitted messages $W_1$ and $W_2$. A rate tuple $(R_1, R_2)$ is said to be achievable, if there exists encoding and decoding functions at the transmitters and the receivers respectively, such that the decoding error probabilities $\lambda_{1,n}, \lambda_{2,n}$ go to zero as $n$ goes to infinity. The capacity region $C(p_d, p_e)$ is the closure of all achievable rate tuples. In the next section, we present the main results of the paper.

| ID | ch. realization | ID | ch. realization | ID | ch. realization | ID | ch. realization |
|----|-----------------|----|-----------------|----|-----------------|----|-----------------|
| 1  | 2               | 3  | 4               | 5  | 6               | 7  | 8               |
| 9  | 10              | 11 | 12              | 13 | 14              | 15 | 16              |

**TABLE I**

ALL POSSIBLE CHANNEL REALIZATIONS; SOLID ARROW FROM TRANSMITTER TX$_i$ TO RECEIVER RX$_j$ INDICATES THAT $G_{ij}[t] = 1.$
III. MAIN RESULTS

The following theorem states our main contribution.

**Theorem 1:** The capacity region of the two-user BFIC with no and delayed local CSIT, $C(p_d, p_c)$, is given by

$$C(p_d, p_c) = \begin{cases} (R_1, R_2) \mid 0 \leq R_i \leq p_d & i = 1, 2 \\ R_i + \beta R_1 \leq \beta p_d + p_c - p_d p_c \end{cases}$$

where

$$\beta = \max \left\{ \frac{p_d - p_c}{p_d p_c}, 1 \right\}.$$  \hspace{1cm} (9)

As mentioned before, the achievability of Theorem [1] is obtained by applying point-to-point erasure codes with appropriate rates at each transmitter, and using either treat-interference-as-erasure or interference-decoding at each receiver, based on the channel parameters. More precisely, for $0 \leq p_c < p_d/(1 + p_d)$ (i.e. $\beta > 1$), the capacity region is obtained by treating interference as erasure, while for $p_d/(1 + p_d) \leq p_c \leq 1$ (i.e. $\beta = 1$), the capacity region is obtained by interference-decoding (i.e. the intersection of the capacity regions of the two multiple access channels at receivers). The detailed proof of achievability can be found in Section [V].

---

**Fig. 2.** (a) Capacity region for $p_d = 1$ and $p_c = 0, 0.4, 0.5$; and (b) capacity region for $p_d = 0.5$ and $p_c = 0, 0.1, 0.5$.

**Fig. 3.** Sum-capacity of the two-user Binary Fading IC with delayed local CSIT, for different values of $p_d$, as $p_c$ moves from 0 to 1.

---

Fig. [2] illustrates the capacity region for several values of $p_d$ and $p_c$. In Fig. [2]a) the capacity region is depicted for $p_d = 1$ and $p_c = 0, 0.4, 0.5$. From [8], we know that

$$C(1, p_c) = C(1, 0.5) \quad \text{for } p_c \in [0.5, 1].$$  \hspace{1cm} (10)
In this case (i.e. \( p_d = 1 \)), considering the maximum achievable sum-rate (or sum-capacity), for small values of \( p_c \), the interference cannot be decoded and as \( p_c \) increases, the sum-capacity decreases as depicted in Fig. 3. However, when \( p_d / (1 + p_d) \leq p_c \), the cross link is strong enough to decode both the intended signal and the interference. We note that compared to the famous “W” curve for the Gaussian interference channel with perfect CSIT\(^{[11]}\), here, we only observe a “V” curve. As mentioned before and depicted in Fig. 3 for \( 0 \leq p_c < p_d / (1 + p_d) \), the capacity region is obtained by treating interference as erasure, while for \( p_d / (1 + p_d) \leq p_c \leq 1 \), the capacity region is obtained by interference-decoding.

From (8) and Fig. 2(b), we conclude that unlike the case in Fig. 2(a), decreasing \( p_c \) does not necessarily enlarge the capacity region. In fact, for \( p_d = 0.5 \), we have

\[
\mathcal{C}(0.5, 0.5) \not\subseteq \mathcal{C}(0.5, 0.1),
\]

\[
\mathcal{C}(0.5, 0.1) \not\subseteq \mathcal{C}(0.5, 0.5).
\]

**Remark 1:** Using Theorem 1 and the results of [12], we have plotted the capacity region of the two-user Binary Fading IC for \( p_d = p_c = 0.5 \) under three different scenarios in Fig. 4. Under the delayed CSIT, we assume that transmitters become aware of all channel realizations with unit delay and receivers have access to instantaneous CSI; and under instantaneous CSIT model, we assume that at time instant \( t \), transmitters and receivers have access to \( G^t \). As we can see in Fig. 4 when transmitters have access to the delayed knowledge of all links in the network, significant capacity gains can be obtained compared to the no CSIT assumption.

To derive the outer-bound, we incorporate two key lemmas as discussed in Section IV-A. The first step in obtaining the outer-bound is to create a “contracted” channel that has fewer states compared to the original channel. Using the Correlation Lemma, we show that an outer-bound on the capacity region of the contracted channel also serves as an outer-bound for the original channel. Finally, using the Conditional Entropy Leakage Lemma, we derive this outer-bound.

Rest of the paper is dedicated to the proof of Theorem 1. We provide the converse proof for the no CSIT case in Section IV, and we present our achievability strategy in Section V. Then in Section VI, we show that same outer-bounds also apply to the case of delayed local CSIT.

**IV. Converse**

In this section, we provide the converse proof of Theorem 1 for the no CSIT assumption. We incorporate two lemmas in order to derive the outer-bound that we describe in the following subsection.

**A. Key Lemmas**

1) **Entropy Leakage Lemma:** Consider a broadcast channel as depicted in Fig. 5 where a transmitter is connected to two receivers through binary fading channels. Suppose \( G_i[t] \) is distributed as i.i.d. Bernoulli random variable (i.e. \( G_i[t] \sim B(p_i) \)) where \( 0 \leq p_2 \leq p_1 \leq 1, \ i = 1, 2 \). In this channel the received signals are given as

\[
Y_i[t] = G_i[t]X[t], \quad i = 1, 2,
\]

\[\text{(12)}\]
where $X[t] \in \{0, 1\}$ is the transmit signal at time instant $t$. Furthermore, suppose $G_3[t] \overset{d}{\sim} \mathcal{B}(p_3)$, and we have
\[
\Pr [G_i[t] = 1, G_j[t] = 1] = 0, \ i \neq j, \ i, j \in \{1, 2, 3\}.
\] (13)

We further define
\[
G_T[t] \overset{d}{=} (G_1[t], G_2[t], G_3[t]) .
\] (14)

Then, for the channel described above, we have the following lemma.

**Lemma 1: [Conditional Entropy Leakage Lemma]** For the channel described above with no CSIT, and for any input distribution, we have
\[
H (Y^n_T | G_3^n X^n, G_T^n) \geq \frac{p_2}{p_1} H (Y^n_1 | G_3^n X^n, G_T^n, G_3^n) .
\] (15)

**Proof:** Let $G_H[t]$ be distributed as $\mathcal{B}(p_2/p_1)$, and be independent of all other parameters in the network. Let
\[
Y_H[t] = G_H[t] Y_1[t], \quad t = 1, \ldots, n .
\] (16)

It is straightforward to see that $Y_H^t$ is statistically the same as $Y_2^t$ under the no CSIT assumption. For time instant $t$, where $1 \leq t \leq n$, we have
\[
\begin{align*}
H (Y^n_2 | Y_2^{t-1}, G_3^n X^n, G_T^n) & \overset{(a)}{=} H (Y^n_2 | Y_2^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}) \\
& \overset{(b)}{=} (1 - p_3) \left[ H (Y^n_1 | Y_2^{t-1}, G_3^n X^n, G_T^n) - H (Y^n_1 | Y_2^{t-1}, G_3^n X^n, G_T^n, G_3^n) \right] \\
& \overset{(c)}{=} p_2 H (X[t] | Y_2^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_2[t] = 1, G_3[t] = 0) \\
& \overset{(d)}{=} p_2 H (X[t] | Y_2^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_3[t] = 0) \\
& \overset{(e)}{=} p_2 H (X[t] | Y_H^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_3[t] = 0) \\
& \overset{(f)}{=} p_2 H (X[t] | Y_1^{t-1}, Y_H^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_3[t] = 0) \\
& \overset{(g)}{=} p_2 H (X[t] | Y_1^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_3[t] = 0) \\
& \overset{(h)}{=} (1 - p_3) \frac{p_2}{p_1} H (Y_1[t] | Y_1^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}, G_3[t] = 0) \\
& \overset{(i)}{=} \frac{p_2}{p_1} H (Y_1[t] | Y_1^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}) \\
& \overset{(j)}{=} \frac{p_2}{p_1} H (Y_1[t] | Y_1^{t-1}, G_3^n X^n, G_T^n) .
\end{align*}
\] (17)

where (a) holds since $G_H^{t-1}$ is independent of all other parameters in the network; (b) is true due to (63); (c) follows from
\[
\Pr [G_2[t] = 1 | G_3[t] = 0] = \frac{p_2}{1 - p_3} ;
\] (18)

(d) holds since the transmit signal is independent of the channel realizations; (e) follows from the fact that $Y_H^{t-1}$ is statistically the same as $Y_2^{t-1}$; (f) holds since conditioning reduces entropy; (g) is true since
\[
H (Y_H^{t-1} | Y_1^{t-1}, G_3^n X^n, G_T^n, G_H^{t-1}) = 0 ;
\] (19)
(h) follows from the fact that
\[
\Pr [G_1[t] = 1 | G_3[t] = 0] = \frac{p_1}{1 - p_3};
\]
(\textit{i}) is true since \( \Pr [G_3[t] = 0] = 1 - p_3 \); and (\textit{j}) holds since \( G_{\theta H}^{t+1} \) is independent of all other parameters in the network. Thus, summing all terms for \( t = 1, \ldots, n \), we get
\[
\sum_{t=1}^{n} H (Y_2[t] | Y_2^t, G_3^n X^n, G_T^n) \geq \frac{p_2}{p_1} \sum_{t=1}^{n} H (Y_1[t] | Y_1^t, G_3^n X^n, G_T^n),
\]
which implies
\[
H (Y_2^n | G_3^n X^n, G_T^n) \geq \frac{p_2}{p_1} H (Y_1^n | G_3^n X^n, G_T^n),
\]
hence, completing the proof.

Remark 2: In [12], we derived the Entropy Leakage Lemma for the case where the transmitter has the CSI with delay. We observe that, with delayed CSIT, the constant on the RHS of (15) would be smaller, meaning that the transmitter can further favor the stronger receiver.

Remark 3: A similar leakage lemma has also been developed in [13], in the case that there are two distributed transmitters employing linear schemes. In this case, a “Rank Ratio Inequality”, is developed that bounds the maximum ratio of the dimensions of received linear-subspaces (at the two receivers) that are created by distributed transmitters with delayed CSIT.

2) Correlation Lemma: Consider again a binary fading interference channel similar to the channel described in Section II but where channel gains have certain correlation. We denote the channel gain from transmitter \( T_{x_i} \) to receiver \( R_{x_j} \) at time instant \( t \) by \( \tilde{G}_{ij}[t] \), \( i, j \in \{1, 2\} \). We distinguish the RVs in this channel, using \( (\cdot) \) notation (e.g., \( \tilde{X}_{1}[t] \)). The input-output relation of this channel at time instant \( t \) is given by
\[
\tilde{Y}_{i}[t] = \tilde{G}_{ii}[t] \tilde{X}_{i}[t] + \tilde{G}_{ri}[t] \tilde{X}_{r}[t], \quad i = 1, 2.
\]
We assume that the channel gains are distributed independent over time. However, they can be arbitrary correlated with each other subject to the following constraints.
\[
\begin{align*}
\Pr \left( \tilde{G}_{ii}[t] = 1 \right) &= p_d, & \Pr \left( \tilde{G}_{ri}[t] = 1 \right) &= p_c, \\
\Pr \left( \tilde{G}_{ii}[t] = 1, \tilde{G}_{ri}[t] = 1 \right) &= \Pr \left( \tilde{G}_{ri}[t] = 1 \right) \Pr \left( \tilde{G}_{ii}[t] = 1 \right),
\end{align*}
\]
In other words, the channel gains corresponding to incoming links at each receiver are still independent. Similar to the original channel, we assume that the transmitters in this BFIC have no knowledge of the CSI. We have the following result.

Lemma 2: [Correlation Lemma] For any BFIC that satisfies the constraints in (24), we have
\[
C (p_d, p_c) \subseteq \tilde{C} (p_d, p_c).
\]
Proof: Suppose in the original BFIC messages \( W_1 \) and \( W_2 \) are encoded as \( X_1^n \) and \( X_2^n \) respectively, and each receiver can decode its corresponding message with arbitrary small decoding error probability as \( n \to \infty \). Now, we show that if we use the same transmission scheme in the BFIC that satisfies the constraints in (24), i.e.
\[
\tilde{X}_{i}[t] = X_{i}[t], \quad t = 1, 2, \ldots, n, \quad i = 1, 2,
\]
then the receivers in this BFIC can still decode \( W_1 \) and \( W_2 \).

In the original two-user binary fading IC as described in Section II, \( R_{x_i} \) uses the decoding function \( \tilde{W}_{i} = \varphi \left( Y_{i}^{n}, \tilde{G}_{ii}^{n}, \tilde{G}_{ii}^{n} \right) \). Therefore, the error event \( \tilde{W}_{i} \neq W_{i} \), only depends on the choice of \( W_{i} \) and marginal distribution of the channel gains \( \tilde{G}_{ii}^{n} \) and \( G_{ii}^{n} \).
Define
\[ \mathcal{E}_W = \left\{ (W_2, G_{12}^n, G_{22}^n) \text{ s.t. } \tilde{W}_1 \neq W_1 \right\}, \]
\[ \tilde{\mathcal{E}}_W = \left\{ (W_2, \tilde{G}_{12}^n, \tilde{G}_{22}^n) \text{ s.t. } \tilde{W}_1 \neq W_1 \right\}, \] (27)
then, the probability of error is given by
\[ p_{\text{error}} = \sum_{w_1} \Pr(W_1 = w_1) \Pr(\mathcal{E}_{w_1}) \]
\[ = \frac{1}{2^n R_1} \sum_{w_1} \sum_{(w_2, \tilde{g}_{12}^n, \tilde{g}_{22}^n) \in \mathcal{E}_{w_1}} \Pr(w_2, \tilde{g}_{12}^n, \tilde{g}_{22}^n) \] (28)
\[ \overset{(a)}{=} \frac{1}{2^n R_1} \sum_{w_1} \sum_{(w_2, \tilde{g}_{12}^n, \tilde{g}_{22}^n) \in \tilde{\mathcal{E}}_{w_1}} \Pr(w_2, \tilde{g}_{12}^n, \tilde{g}_{22}^n) = \tilde{p}_{\text{error}}, \]
where \( p_{\text{error}} \) and \( \tilde{p}_{\text{error}} \) are the decoding error probability at \( Rx_1 \) in the original and the BFIC satisfying the constraints in (24) respectively; and \((a)\) holds since according to (24), the joint distribution of \( G_{12}^n \) and \( G_{22}^n \) is the same as \( \tilde{G}_{12}^n \) and \( \tilde{G}_{22}^n \) and the fact that, as mentioned above, the error probability at receiver one only depends on the marginal distribution of these links. Similar argument holds for \( Rx_2 \).

B. Deriving the Outer-bounds

The key to derive the outer-bound is the proper application of the two lemmas introduced in Section IV-A. More precisely, we need to find a channel that satisfies the constraints in (24) such that the outer-bound on its capacity region, coincides with the achievable region of the original problem. We provide the proof for three separate regimes.

• **Regime I:** \( 0 \leq p_c \leq \frac{p_d}{1 + p_d} \): The derivation of the individual bounds, e.g., \( R_1 \leq p_d \), is straightforward and omitted here. We focus on

\[ R_1 + \beta R_2 \leq \beta p_d + p_c - p_d p_c, \] (29)

where

\[ \beta = \max \left\{ \frac{p_d - p_c}{p_d p_c}, 1 \right\}. \] (30)

Due to symmetry, the derivation of the other bound is similar.

### TABLE II
**The Contracted Channel for Regime I and Regime II.**

| ID | channel realization | ID | channel realization |
|----|---------------------|----|---------------------|
| A  | \( \tau_e \) \( \tau_e \) \( \tau_e \) | B  | \( \tau_e \) \( \tau_e \) |
|    | \( \tau_e \) \( \tau_e \) \( \tau_e \) |    | \( \tau_e \) \( \tau_e \) |
| Pr [state A] = \( p_d p_c \) | Pr [state B] = \( p_d - p_c \) |
| C  | \( \tau_e \) \( \tau_e \) \( \tau_e \) | D  | \( \tau_e \) \( \tau_e \) |
|    | \( \tau_e \) \( \tau_e \) \( \tau_e \) |    | \( \tau_e \) \( \tau_e \) |
| Pr [state C] = \( q_d p_c \) | Pr [state D] = \( q_d p_c \) |

The first step is to define the appropriate channel that satisfies the constraints in (24). The idea is to construct a channel such that \( \tilde{G}_{ij}[t] = 1 \) whenever \( \tilde{G}_{ii}[t] = 1, i = 1, 2 \). We construct such channel with only five states rather than 16 states and thus, we refer to it as “contracted” channel. The five states are denoted by states \( A, B, C, D, \)
and $E$ with corresponding probabilities $p_d p_c, (p_d - p_c), q_d p_c, q_d p_c$, and $q_d q_c$. These states are depicted in Table II with the exception of state $E$ which corresponds to the case where all channel gains are 0. Here, we have

$$
\Pr \left( G_{11}[t] = 1 \right) = \sum_{j \in \{A,B,C\}} \Pr (\text{State } j) = p_d,
$$

$$
\Pr \left( G_{12}[t] = 1 \right) = \sum_{j \in \{A,C\}} \Pr (\text{State } j) = p_c,
$$

$$
\Pr \left( G_{11}[t] = 1, G_{21}[t] = 1 \right) = \Pr (\text{State } A) = p_d p_c,
$$

thus, this channel satisfies the conditions in (24). From Lemma 2, we have $\mathcal{C}(p_d, p_c) \subseteq \mathcal{C}(p_d, p_c)$, thus, any outer-bound on the capacity region of the contracted channel, provides an outer-bound on the capacity region of the original channel.

We define

$$
\hat{X}_{1A}[t] \triangleq \tilde{X}_1[t] 1_{\{\text{state } A \text{ occurs at time } t\}},
$$

where $1_{\{\text{state } A \text{ occurs at time } t\}}$ is equal to 1 when at time $t$ state $A$ occurs. Similarly, we define $\tilde{X}_{1B}[t], \tilde{X}_{1C}[t], \tilde{X}_{1D}[t], \tilde{X}_{1E}[t], \tilde{X}_{2A}[t], \tilde{X}_{2B}[t], \tilde{X}_{2C}[t], \tilde{X}_{2D}[t]$ and $\tilde{X}_{2E}[t]$. Therefore, we have

$$
\tilde{X}_1[t] = \tilde{X}_{1A}[t] \oplus \tilde{X}_{1B}[t] \oplus \tilde{X}_{1C}[t] \oplus \tilde{X}_{1D}[t] \oplus \tilde{X}_{1E}[t].
$$

Suppose in the contracted channel, there exist encoders and decoders at transmitters and receivers respectively, such that each receiver can decode its corresponding message with arbitrary small decoding error probability as $\epsilon_n \to 0$. The derivation of the outer-bound is given in (33) for $\epsilon_n \to 0$ as $n \to \infty$; and

$$
0 \leq p_c \leq p_d / (1 + p_d) \implies \beta = \frac{p_d - p_c}{p_d p_c} > 1.
$$

We have

$$
n \left( \tilde{R}_1 + \beta \tilde{R}_2 - \epsilon_n \right) \overset{(a)}{\leq} I \left( \tilde{X}_1^n; \tilde{Y}_1^n | \tilde{G}^n \right) + \beta I \left( \tilde{X}_2^n; \tilde{Y}_2^n | \tilde{G}^n \right) \overset{(b)}{=} H \left( \tilde{X}_n^{2d} | \tilde{G}^n \right) + H \left( \tilde{X}_n^{1B} | \tilde{X}_n^{1C}, \tilde{X}_n^{2b}, \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{2A}, \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{X}_n^{2D}, \tilde{G}^n \right) \overset{(c)}{\leq} H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1C}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1B} | \tilde{X}_n^{1C}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{2A} | \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{2A}, \tilde{X}_n^{1C}, \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1B}, \tilde{X}_n^{1C}, \tilde{X}_n^{2A}, \tilde{X}_n^{2D}, \tilde{G}^n \right) \overset{(d)}{=} H \left( \tilde{X}_n^{1C}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{1B} | \tilde{X}_n^{1C}, \tilde{G}^n \right) + (1 + \beta) H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{2A}, \tilde{G}^n \right) - H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{1C}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{2B} | \tilde{X}_n^{2D}, \tilde{G}^n \right) - H \left( \tilde{X}_n^{2A} | \tilde{X}_n^{2D}, \tilde{G}^n \right) \overset{(e)}{\leq} H \left( \tilde{X}_n^{1C}, \tilde{G}^n \right) + (1 + \beta) H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{2A}, \tilde{G}^n \right) + \beta H \left( \tilde{X}_n^{2B} | \tilde{X}_n^{2D}, \tilde{G}^n \right) - H \left( \tilde{X}_n^{2A} | \tilde{X}_n^{2D}, \tilde{G}^n \right) \overset{(f)}{\leq} H \left( \tilde{X}_n^{1C}, \tilde{G}^n \right) + (1 + \beta) H \left( \tilde{X}_n^{1A} | \tilde{X}_n^{2A}, \tilde{G}^n \right) + \beta H \left( \tilde{X}_n^{2B} | \tilde{X}_n^{2D}, \tilde{G}^n \right) + H \left( \tilde{X}_n^{2B} | \tilde{X}_n^{2D}, \tilde{G}^n \right) \overset{(g)}{\leq} n q_d p_c + n (1 + \beta) p_d p_c + n \beta q_d p_c + n \left( \beta - \frac{1}{\beta} \right) \left( p_d - p_c \right) = n (\beta p_d + p_c - p_d p_c). \quad (35)
where \((a)\) follows from Fano’s inequality and data processing inequality; \((b)\) follows from the definition of the contracted channel and the chain rule; \((c)\) is true since from Claim [1] below, we have

\[
I \left( \bar{X}_1^n; \bar{X}_2^n | \bar{G}^n \right) = 0,
\]

which results in

\[
H \left( \bar{X}_2^n | \bar{X}_1^n, \bar{G}_n \right) = H \left( \bar{X}_2^n | \bar{X}_2^n, \bar{G}_n \right),
\]

\[
H \left( \bar{X}_1^n | \bar{X}_1^n, \bar{G}_n \right) = H \left( \bar{X}_1^n | \bar{X}_1^n, \bar{G}_n \right);
\]

and the fact that conditioning reduces entropy; \((d)\) follows from the chain rule; \((e)\) holds since using Lemma [1] we have

\[
H \left( \bar{X}_1^n, \bar{X}_1^n, \bar{G}_n \right) - \beta H \left( \bar{X}_1^n | \bar{X}_1^n, \bar{G}_n \right) \leq 0;
\]

and \((f)\) follows since from applying Lemma [1] we have

\[
H \left( \bar{X}_2^n | \bar{X}_2^n, \bar{G}_n \right) \geq \frac{1}{\beta} H \left( \bar{X}_2^n | \bar{X}_2^n, \bar{G}_n \right);
\]

and \((g)\) follows from the fact that entropy of a binary random variable is maximized by i.i.d. Bernoulli distribution with success probability of half. Dividing both sides by \(n\) and let \(n \to \infty\), we get the desired result.

**Claim 1:**

\[
I \left( \bar{X}_1^n; \bar{X}_2^n | \bar{G}^n \right) = 0.
\]

**Proof:**

\[
0 \leq I \left( \bar{X}_1^n; \bar{X}_2^n | \bar{G}^n \right) \leq I \left( \bar{W}_1, \bar{X}_1^n; \bar{W}_2, \bar{X}_2^n | \bar{G}^n \right).
\]

\[
= I \left( \bar{W}_1 \bar{W}_2 | \bar{G}^n \right) + I \left( \bar{W}_1 \bar{X}_2 | \bar{W}_2, \bar{G}^n \right) + I \left( \bar{X}_2^n | \bar{W}_2, \bar{X}_2^n | \bar{W}_2, \bar{G}^n \right) = 0.
\]

\[
= 0 \text{ since } \bar{W}_1 \perp \bar{W}_2 | \bar{G}^n = 0 \text{ since } \bar{X}_2^n = \bar{f}_2(\bar{W}_2, \bar{G}^n) = 0 \text{ since } \bar{X}_1^n = \bar{f}_1(\bar{W}_1, \bar{G}^n)
\]

**• Regime II:** \(p_d/ (1 + p_d) \leq p_c \leq p_d\): In this regime, the capacity region would be equal to the intersection of capacity regions of the two multiple-access channels (MACs) formed at the receivers. Thus, we need to show that

\[
R_1 + R_2 \leq p_d + p_c - p_d p_c.
\]

We use the same contracted channel as for the case of **Regime I**. Let \(G_S[t]\) be distributed as i.i.d. Bernoulli RV and

\[
G_S[t] \overset{d}{\sim} \mathcal{B}(\frac{p_d - p_c}{p_d p_c}),
\]

and for \(i = 1, 2\), we define

\[
\bar{X}_{iA}[t] \overset{\Delta}{=} G_S[t] \bar{X}_{iA}[t],
\]

\[
\bar{X}_{iA}[t] \overset{\Delta}{=} (1 - G_S[t]) \bar{X}_{iA}[t],
\]

\[
(44)
\]
We have
\[ n \left( \hat{R}_1 + \hat{R}_2 - \epsilon_n \right) \leq I \left( \hat{X}_1^n, \hat{Y}_1^n \mid \hat{G}^n \right) + I \left( \hat{X}_2^n \mid \hat{Y}_2^n \mid \hat{G}^n \right) \]
\[ = \text{(a)} \quad H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A}, \hat{X}_{1B}, \hat{X}_{2B}, \hat{X}_{1C}, \hat{X}_{2D} \mid \hat{G}^n \right) - H \left( \hat{X}_{2A}, \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D} \mid \hat{G}^n \right) \]
\[ + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A}, \hat{X}_{1C}, \hat{X}_{2D} \mid \hat{G}^n \right) - H \left( \hat{X}_{1A}, \hat{X}_{1C} \mid \hat{G}^n \right) \]
\[ = \text{(b)} \quad H \left( \hat{X}_{2D} \mid \hat{G}^n \right) + H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{X}_{2D}, \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ - H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{X}_{2D}, \hat{G}^n \right) \]
\[ = \text{(c)} \quad H \left( \hat{X}_{2D} \mid \hat{G}^n \right) + H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{X}_{2D}, \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ - H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{X}_{2D}, \hat{G}^n \right) \]
\[ = \text{(d)} \quad H \left( \hat{X}_{2D} \mid \hat{G}^n \right) + H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{X}_{2D}, \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ - H \left( \hat{X}_{2D} \mid \hat{G}^n \right) \]
where \( \epsilon_n \to 0 \) as \( n \to \infty \); (a) follows from Fano’s inequality and data processing inequality; (b) holds due to the definition of the contracted channel; (c) follows from the chain rule; (d) is true since from Claim \[ \square \] we have
\[ I \left( \hat{X}_1^n, \hat{X}_2^n \mid \hat{G}^n \right) = 0. \]

We use our definition in (44) for the rest of the proof.
\[ = \text{(e)} \quad H \left( \hat{X}_{1A} \mid \hat{G}^n \right) + H \left( \hat{X}_{1B}, \hat{X}_{1C} \mid \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ - H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ \leq \text{(f)} \quad H \left( \hat{X}_{1C} \mid \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \bigoplus \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ \leq \text{(g)} \quad H \left( \hat{X}_{1C} \mid \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \bigoplus \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ \leq \text{(h)} \quad H \left( \hat{X}_{1C} \mid \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \bigoplus \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ \leq \text{(i)} \quad H \left( \hat{X}_{1C} \mid \hat{G}^n \right) + H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \bigoplus \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) \]
\[ \leq \text{(j)} \quad n \left( p_d + p_c - p_d p_c + \epsilon_n \right), \]
\[ \text{where (e) follows from (44) and the fact that } \hat{X}_{1A} \text{ and } \hat{X}_{1B} \text{ are statistically the same (the reason is that } X_{1A}[t] \text{ and } X_{1B}[t] \text{ only depend on } W_1 \text{ and } G_{11}^{t-1}, \text{ thus,} \]
\[ H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \bigoplus \hat{X}_{1B}, \hat{X}_{1C}, \hat{X}_{2D}, \hat{G}^n \right) = H \left( \hat{X}_{1A} \bigoplus \hat{X}_{2A} \mid \hat{X}_{1B}, \hat{X}_{1C}, \hat{G}^n \right), \]
and similar statement is true for $\tilde{X}_{1A}^n$ and $\tilde{X}_{2B}^n$; (f) holds since from Lemma 1 we have

$$H(\tilde{X}_{1B}^n|\tilde{X}_{1C}^n, \hat{G}^n) - H(\tilde{X}_{1A}^n|\tilde{X}_{1C}^n, \tilde{G}^n) \leq 0;$$  \hspace{1cm} (48)

$$(g)$$ holds since

$$H(\tilde{X}_{1A}^n|\tilde{X}_{1A}^n \oplus \tilde{X}_{2A}^n, \tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{X}_{2D}^n, \hat{G}^n) - H(\tilde{X}_{1A}^n|\tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{G}^n) \leq 0;$$  \hspace{1cm} (49)

$h$ is true since we have

$$H(W_1|\tilde{X}_{2A}^n, \tilde{X}_{1A}^n \oplus \tilde{X}_{2A}^n, \tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{X}_{2D}^n, \hat{G}^n) \leq n\epsilon_n,$$  \hspace{1cm} (50)

and as a result, we have

$$H(\tilde{X}_{2A}^n, \tilde{X}_{1A}^n, \tilde{X}_{1A}^n \oplus \tilde{X}_{2A}^n, \tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{X}_{2D}^n, \tilde{G}^n) \leq H(W_1) + H(\tilde{X}_{2A}^n, \tilde{X}_{2D}^n|\tilde{G}^n) + H(\tilde{X}_{1A}^n|\tilde{X}_{1A}^n, \tilde{X}_{2D}^n, \tilde{G}^n) + n\epsilon_n,$$  \hspace{1cm} (51)

we then use the fact that $\tilde{X}_{2A}^n$ and $\tilde{X}_{2B}^n$ are statistically the same; (i) holds since $\tilde{X}_{2A}^n$ and $\tilde{X}_{2B}^n$ are statistically the same; and (j) holds since

$$H(\tilde{X}_{1C}^n|\tilde{G}^n) \leq nq_d p_c,$$

$$H(\tilde{X}_{2D}^n|\tilde{G}^n) \leq nq_d p_c,$$

$$H(\tilde{X}_{1A}^n \oplus \tilde{X}_{2A}^n|\tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{X}_{2D}^n, \tilde{G}^n) \leq np_d p_c,$$

$$H(\tilde{X}_{1A}^n \oplus \tilde{X}_{2A}^n|\tilde{X}_{1B}^n, \tilde{X}_{1C}^n, \tilde{X}_{2D}^n, \tilde{G}^n) \leq n(p_d - p_c).$$  \hspace{1cm} (52)

**Regime III:** $p_d \leq p_c \leq 1$: In this regime, again we have $\beta = 1$, and the capacity region would be equal to the intersection of capacity regions of the two MACs formed at the receivers. Under no CSIT assumption, we do not need to create a contracted channel. However, since we need a contracted channel for the delayed local CSIT assumption, we present the proof using a contracted channel. We consider a contracted channel with five states denoted by $A, B, C, D,$ and $E$ with corresponding probabilities $p_d p_c, (p_c - p_d), p_d q_c, p_d q_c$, and $q_d q_c$. These states are depicted in Table III with the exception of state $E$ which corresponds to the case where all channel gains are 0. We assume that $\tilde{G}^n_{ii} = G^n_{ii}$. Here, we have

$$\Pr(\tilde{G}_{11}[t] = 1) = \sum_{j \in \{A, C\}} \Pr(\text{State } j) = p_d,$$

$$\Pr(\tilde{G}_{12}[t] = 1) = \sum_{j \in \{A, B, C\}} \Pr(\text{State } j) = p_c,$$

$$\Pr(\tilde{G}_{11}[t] = 1, \tilde{G}_{21}[t] = 1) = \Pr(\text{State } A) = p_d p_c,$$  \hspace{1cm} (53)

thus, this channel satisfies the conditions in (24). Since from Lemma 2 we have $C(p_d, p_c) \subseteq \tilde{C}(p_d, p_c)$, any outer-bound on the capacity region of the contracted channel, provides an outer-bound on the capacity region of the original problem.

The derivation of the outer-bound is easier compared to the other regimes. Basically, $R_{X_i}$ after decoding and removing its corresponding signal, has a stronger channel from $T_{X_i}$ compared to $R_{X_i}$, and thus it must be able to decode both $\tilde{W}_i$ and $\tilde{W}_i^\dagger$, $i = 1, 2$. Thus, we have

$$R_1 + R_2 \leq p_d + p_c - p_d p_c.$$  \hspace{1cm} (54)
TABLE III
THE CONTRACTED CHANNEL FOR Regime III.

| ID | channel realization | ID | channel realization |
|----|---------------------|----|---------------------|
| A  | ![Figure A]          | B  | ![Figure B]          |
|    | \( \Pr [\text{state } A] = p_d p_c \) |    | \( \Pr [\text{state } B] = p_c - p_d \) |
| C  | ![Figure C]          | D  | ![Figure D]          |
|    | \( \Pr [\text{state } C] = p_d q_c \) |    | \( \Pr [\text{state } D] = p_d q_c \) |

V. ACHIEVABILITY

In this section, we provide the achievability proof for Theorem 1. We show that the entire capacity region is achieved by applying point-to-point erasure codes with appropriate rates at each transmitter, using either treat-interference-as-erasure or interference-decoding at each receiver, based on the channel parameters.

Each transmitter applies a point-to-point erasure random code as described in [14]. On the other hand, at each receiver we have two options: (1) interference-decoding; and (2) treat-interference-as-erasure. When a receiver decodes the interference alongside its intended message, the achievable rate region is the capacity region of the multiple-access channel (MAC) formed at that receiver as depicted in Fig. 6. The MAC capacity at \( \text{Rx}_1 \), is given by

\[
\begin{cases}
R_1 & \leq p_d, \\
R_2 & \leq p_c, \\
R_1 + R_2 & \leq 1 - q_d q_c.
\end{cases}
\]  

(55)

As a result, \( \text{Rx}_1 \) can decode its message by interference decoding, if \( R_1 \) and \( R_2 \) satisfy the constraints in (55). On the other hand, if \( \text{Rx}_1 \) treats interference as erasure, it basically ignores the received signal at time instants where \( G_{21}[t] = 1 \), and \( p_d q_c \) fraction of the time, it receives the transmit signal of \( \text{Tx}_1 \). Thus, \( \text{Rx}_1 \) can decode its message by treat-interference-as-erasure, if \( R_1 \leq p_d q_c \).

Similarly, \( \text{Rx}_2 \) can decode its message by treat-interference-as-erasure, if \( R_2 \leq p_d q_c \). Also, \( \text{Rx}_2 \) can decode its message by interference decoding, if \( R_1 \) and \( R_2 \) are inside the capacity region of the multiple-access channel (MAC) formed at \( \text{Rx}_2 \), i.e.,

\[
\begin{cases}
R_1 & \leq p_c, \\
R_2 & \leq p_d, \\
R_1 + R_2 & \leq 1 - q_d q_c.
\end{cases}
\]  

(56)

Therefore, the achievable rate region by either treat-interference-as-erasure or interference-decoding at each
receiver is the convex hull of

\[ \mathcal{R} = \left\{ (R_1, R_2) \left| \begin{array}{l}
R_1 \leq p_d \\
R_2 \leq p_c \\
R_1 + R_2 \leq 1 - q_d q_c
\end{array} \right\} \text{ or } R_1 \leq p_d q_c \right\} \text{ and } \left\{ (R_1, R_2) \left| \begin{array}{l}
R_1 \leq p_c \\
R_2 \leq p_d \\
R_1 + R_2 \leq 1 - q_d q_c
\end{array} \right\} \right\} \]

(57)

In the remaining of this section, we show that the convex hull of \( \mathcal{R} \) matches the outer-bound of Theorem 1, i.e.

\[ \left\{ (R_1, R_2) \left| 0 \leq R_i \leq p_d \right. \right\} \text{ or } R_i + \beta R_i \leq \beta p_d + p_c - p_d p_c \]

(58)

where

\[ \beta = \max \left\{ \frac{p_d - p_c}{p_d p_c}, 1 \right\} \]

(59)

First, it is easy to verify that the region described in (58), is the convex hull of the following corner points:

- \( (R_1, R_2) = (0, 0) \),
- \( (R_1, R_2) = (p_d, 0) \),
- \( (R_1, R_2) = (0, p_d) \),
- \( (R_1, R_2) = (p_d, q_d p_c) \),
- \( (R_1, R_2) = (q_d p_c, p_d) \),
- \( (R_1, R_2) = (p_d q_c, p_d q_c) \).

(60)

Now, when \( 0 \leq p_c \leq p_d/(1 + p_d) \), the rate region \( \mathcal{R} \) (as defined in (57)) and its convex hull are shown in Fig. 7. As we can note from Fig. 7(b), in this case the convex hull of \( \mathcal{R} \) is indeed the convex hull of the corner points in (60).

Fig. 7. (a) Depiction of the rate region \( \mathcal{R} \) for \( 0 \leq p_c \leq p_d/(1 + p_d) \); and (b) its convex hull.

On the other hand, when \( p_d/(1 + p_d) < p_c \leq 1 \), the rate region \( \mathcal{R} \) (as defined in (57)) is depicted in Fig. 8, which is the convex hull of the first five corner points in (60). In this case, the last point in (60) is strictly inside the region in Fig. 8 hence again, \( \mathcal{R} \) coincides with the convex hull of the corner points in (60), and this completes the proof of Theorem 1 under no CSIT assumption.
VI. Extension to Delayed Local CSIT

In this section, we complete the proof of Theorem 1 by showing that under delayed local CSIT assumption, the capacity region remains unchanged. We show that same outer-bounds hold under the delayed local CSIT assumption, and as a result, the achievability strategy of Section V suffices.

To prove this result, we first state counterparts of Lemma 1 and Lemma 2 considering the delayed local CSIT. The converse proof then follows exact same step as what we described in Section IV-B, by replacing Lemma 1 and Lemma 2 with Lemma 3 and Lemma 4 respectively.

1) Entropy Leakage Lemma with Delayed Local CSIT: Consider the scenario where a transmitter is connected to two receivers through binary fading channels as in Figure 9. In this channel the received signals are given as

\[ Y_i[t] = G_i[t]X[t], \quad i = 1, 2, \tag{61} \]

where \( X[t] \in \{0, 1\} \) is the transmit signal at time instant \( t \). We assume that at each time instant \( t \), one of the three states given in Table IV may occur. In other words, \( G_2[t] = 1 \) implies \( G_1[t] = 1 \). These states are labeled \( A, B, \) and \( C \), with probabilities \( p_A, p_B, \) and \( (1 - p_A - p_B) \) respectively (\( p_A + p_B \leq 1 \)). We further assume that \( p_A \leq p_B \).

![Fig. 9. A transmitter connected to two receivers through binary fading channels.](image)

We define

\[ X_A[t] \overset{\Delta}{=} X[t]1\{\text{state } A \text{ occurs at time } t\}, \tag{62} \]

and similarly, we define \( X_B[t] \) and \( X_C[t] \). Finally, suppose \( G_3[t] \overset{d}{\sim} B(p_3) \) where \( p_A + p_B + p_3 \leq 1 \), and

\[ \Pr \left[ G_1[t] = 1, G_3[t] = 1 \right] = 0. \tag{63} \]

Set

\[ G_T[t] \overset{\Delta}{=} (G_1[t], G_2[t], G_3[t]). \tag{64} \]

Then, for the channel described above, we have the following lemma. The proof of this lemma is presented in Appendix A.
Lemma 3: [Conditional Entropy Leakage with Delayed Local CSIT] For the channel described above where at time instant $t$, the transmitter has only access to $G_{1}^{d-1}$, and for any input distribution, we have

$$H(X_{A}^{n}|G_{3}^{m}X_{n},G_{T}^{m}) \geq \frac{PA}{PB}H(X_{B}^{n}|G_{3}^{m}X_{n},G_{T}^{m}).$$

(65)

2) Correlation Lemma with Delayed Local CSIT: Consider again a binary fading interference channel similar to the channel described in Section II but where channel gains have certain correlation. We denote the channel gain from transmitter $Tx_i$ to receiver $Rx_j$ at time instant $t$ by $\tilde{G}_{ij}[t]$, $i,j \in \{1,2\}$. We distinguish the RVs in this channel, using $\tilde{\cdot}$ notation (e.g., $\tilde{X}_{1}[t]$). In this new channel, the channel gains are distributed independent over time, however, they can be arbitrary correlated as long as

$$\tilde{G}_{ii} = G_{ii}^{n}, \ i = 1,2,$$

(66)

and

$$Pr(\tilde{G}_{ii}[t] = 1) = p_{c},$$

$$Pr(\tilde{G}_{ii}[t] = 1,\tilde{G}_{ij}[t] = 1) = Pr(\tilde{G}_{ii}[t] = 1)Pr(\tilde{G}_{ij}[t] = 1), \ i = 1,2.$$  (67)

In other words, the channel gains corresponding to incoming links at each receiver are still independent. The input-output relation of the contracted channel at time instant $t$ is given by

$$\tilde{Y}_{i}[t] = \tilde{G}_{ii}[t]\tilde{X}_{i}[t] \oplus \tilde{G}_{ij}[t]\tilde{X}_{j}[t], \ i = 1,2.$$  (68)

We assume that the transmitters have access to delayed local knowledge of the channel state information. Similar to the original BFIC, the capacity region for the channel that satisfies (66) and (67) is denoted by $\tilde{C}(p_{d},p_{c})$. We have the following lemma and its proof is presented in Appendix B.

Lemma 4: [Correlation Lemma with Delayed Local CSIT] Consider the two user BFIC with parameters $p_{d}$ and $p_{c}$ as described in Section II and capacity $C(p_{d},p_{c})$. For any BFIC that satisfies (66) and (67) with capacity $\tilde{C}(p_{d},p_{c})$, we have

$$C(p_{d},p_{c}) \subseteq \tilde{C}(p_{d},p_{c}).$$  (69)

As mentioned before, the converse proof of Theorem 1 under delayed local CSIT assumption follows exact same step as what we described in Section IV-B by replacing Lemma 1 and Lemma 2 with Lemma 3 and Lemma 4 respectively.

VII. CONCLUSION

We characterized the capacity region of the two-user Binary Fading Interference Channel with no CSIT and delayed local. We showed that the entire capacity region is achieved by applying point-to-point erasure codes with appropriate rates at each transmitter, using either treat-interference-as-erasure or interference-decoding at each receiver, based on the channel parameters. We obtained a novel outer-bound that relies on two key lemmas, and we showed that delayed local knowledge of the channel state information, does not enlarge the capacity region.

As mentioned before, the capacity region of the two-user fading Gaussian interference channel with no CSIT remains open. In this work, we considered the two-user Binary Fading Interference Channel with no CSIT, and we observed that in order to achieve the capacity region, message splitting (such as Han-Kobayashi scheme) is not required. One important question then is whether message splitting is required for the two-user fading Gaussian interference channel with no CSIT or not. To answer this question, one path would be to extend the present results to the linear deterministic model (a layered extension of our work, similar to that of [10]) and then, Gaussian model.
The work of A. S. Avestimehr and A. Vahid is in part supported by NSF Grants CAREER-0953117, CCF-1161720, NETS-1161904, AFOSR Young Investigator Program Award, ONR award N000141310094, and 2013 Qualcomm Innovation Fellowship.

APPENDIX A

PROOF OF LEMMA 3

Let $G_{1t}$ be distributed as $B(p_A/p_B)$, and be independent of all other parameters in the network. Let

$$X_{ht} = G_{ht}X_B[t], \quad t = 1, \ldots, n. \quad (70)$$

We note that $X_h^t$ is statistically the same as $X_A^t$, $t = 1, \ldots, n$; and under the assumptions of the lemma (i.e. the transmitter has access to $G_{h}^{t-1}$ at time instant $t$), the two signals are indistinguishable for the transmitter.

For time instant $t$ where $1 \leq t \leq n$, we have

$$H(X_A[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n})$$

$$\leq \begin{align*}
(a) & H(X_A[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}) \\
(b) & (1-p_{A})H(X_A[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{3}[t] = 0) \\
& = p_{A}H(X[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{2}[t] = 1, G_{3}[t] = 0) \\
(c) & p_{A}H(X[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0) \\
(d) & \geq p_{A}H(X[t]|X_{B}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0) \\
(e) & (1-p_{A})\frac{P_{A}}{P_{B}}H(X_B[t]|X_{B}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}) \\
& = \frac{P_{A}}{P_{B}}H(X_B[t]|X_{B}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}) \\
\end{align*}$$

where (a) holds since $G_{h}^{t-1}$ is independent of all other parameters in the network; (b) is true due to (63); (c) follows since due to delayed local CSIT assumption, given the realization of $G_{1}[t]$, the signal is independent of the realization of $G_{2}[t]$; (d) holds since

$$H(X[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0)$$

$$= H(X[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0) - H(X_{A}^{t-1}|G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0)$$

$$= H(X[t]|X_{h}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0) - H(X_{h}^{t-1}|G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0)$$

$$= H(X[t]|X_{h}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}, G_{h}^{t-1}, G_{1}[t] = 1, G_{3}[t] = 0), \quad (72)$$

where the second equality holds since $X[t], X_{A}^{t-1}$ have the same conditional joint distribution as $X[t], X_{h}^{t-1}$, and $X_{h}^{t-1}$ has the same conditional joint distribution as $X_{h}^{t-1}$; (e) is true since from (70), we know that $X_{h}^{t-1}$ is a function of $X_{B}^{t-1}$ and $G_{h}^{t-1}$; (f) follows from the fact that given $G_{3}[t] = 0$ state $B$ occurs with probability $p_{B}/(1-p_{A})$; and (g) holds since $G_{h}^{t-1}$ is independent of all other parameters in the network.

Finally, using (71), we have

$$\sum_{t=1}^{n} H(X_A[t]|X_{A}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}) \geq \frac{P_{A}}{P_{B}} \sum_{t=1}^{n} H(X_B[t]|X_{B}^{t-1}, G_{3}^{n}X_{n}, G_{T}^{n}), \quad (73)$$

which implies

$$H(X_{A}^{n}|G_{3}^{n}X_{n}, G_{T}^{n}) \geq \frac{P_{A}}{P_{B}} H(X_{B}^{n}|G_{3}^{n}X_{n}, G_{T}^{n}), \quad (74)$$

hence, we get the desired result.
APPENDIX B
PROOF OF LEMMA [3]
Suppose messages $W_1$ and $W_2$ are encoded as $X_i^n$ and $X_2^n$ respectively, and each receiver in the original BFIC can decode its corresponding message with arbitrary small decoding error probability as $n \to \infty$.

Now, we show that if we use the same transmission scheme in the BFIC that satisfies (66) and (67), \textit{i.e.}
\begin{equation}
\tilde{X}_i[t] = X_i[t], \quad t = 1, 2, \ldots, n, \quad i = 1, 2,
\end{equation}
then the receivers in the BFIC that satisfies (66) and (67) can still decode $W_1$ and $W_2$.

Note that $X_i[t]$ only depends on $W_i$ and $G^n_{ii}$, $i = 1, 2$. Therefore, the error event $\tilde{W}_i \neq W_i$, depends on the choice of $W_i$, $G^n_{ii}$, and the marginal distribution of the channel gains $G^n_{ii}$ and $G^n_{22}$.

Define
\begin{align*}
\tilde{E}_W &= \{ (W_2, G^n_{12}, G^n_{22}) \text{ s.t. } \tilde{W}_1 \neq W_1 \}, \\
\tilde{E}_W &= \{ (W_2, G^n_{12}, G^n_{22}) \text{ s.t. } \tilde{W}_1 \neq W_1 \},
\end{align*}
then, the probability of error is given by
\begin{equation}
\begin{aligned}
\tilde{p}_{\text{error}} &= \sum_{w_1} \Pr (W_1 = w_1) \Pr (\tilde{E}_W) \\
&= \frac{1}{2^n R_1} \sum_{w_1} \sum_{(w_2, g^n_{12}, g^n_{22}) \in \tilde{E}_W} \Pr (w_2, g^n_{12}, g^n_{22}) \\
&= \frac{1}{2^n R_1} \sum_{w_1} \sum_{(w_2, g^n_{12}, g^n_{22}) \in \tilde{E}_W} \Pr (w_2, g^n_{12}, g^n_{22})
\end{aligned}
\end{equation}
where $\tilde{p}_{\text{error}}$ and $\tilde{p}_{\text{error}}$ are the decoding error probability at $R_{x_1}$ in the original and the BFIC that satisfies (66) and (67) respectively; and (a) holds since according to (24), the joint distribution of $G^n_{12}$ and $G^n_{22}$ is the same as $\tilde{G}^n_{12}$ and $\tilde{G}^n_{22}$ and the fact that the error probability at receiver one only depends on the marginal distribution of these links. Similar argument holds for $R_{x_2}$.

REFERENCES
[1] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Binary Fading Interference Channel with No CSIT,” to be presented in International Symposium on Information Theory (ISIT), 2014.
[2] R. Ahlswede, “The capacity region of a channel with two senders and two receivers,” The Annals of Probability, pp. 805–814, 1974.
[3] T. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 49–60, 1981.
[4] H. Sato, “Two-user communication channels,” IEEE Transactions on Information Theory, vol. 23, no. 3, pp. 295–304, 1977.
[5] A. El-Gamal and M. H. Costa, “The capacity region of a class of deterministic interference channels,” IEEE Transactions on Information Theory, vol. 28, pp. 343–346, Mar. 1982.
[6] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Communication through collisions: Opportunistic utilization of past receptions,” to appear in IEEE Infocom 2014. arXiv preprint arXiv:1312.0116, 2013.
[7] V. Aggarwal, L. Sankar, A. R. Calderbank, and H. V. Poor, “Ergodic layered erasure one-sided interference channels,” in IEEE Information Theory Workshop, pp. 574–578, IEEE, 2009.
[8] Y. Zhu and D. Guo, “Ergodic fading Z-interference channels without state information at transmitters,” IEEE Transactions on Information Theory, vol. 57, no. 5, pp. 2627–2647, 2011.
[9] A. S. Avestimehr, S. Diggavi, and D. Tse, “Wireless network information flow: A deterministic approach,” IEEE Transactions on Information Theory, vol. 57, Apr. 2011.
[10] D. N. Tse and R. D. Yates, “Fading broadcast channels with state information at the receivers,” IEEE Transactions on Information Theory, vol. 58, no. 6, pp. 3453–3471, 2012.
[11] R. Etkin, D. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” IEEE Transactions on Information Theory, vol. 54, pp. 5534–5562, Dec. 2008.
[12] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Capacity results for binary fading interference channels with delayed CSIT,” arXiv preprint arXiv:1301.5309, 2013.
[13] S. Lashgari, A. S. Avestimehr, and C. Suh, “Linear degrees of freedom of the X-Channel with delayed CSIT,” IEEE Transactions on Information Theory, vol. 60, no. 4, 2014.
[14] P. Elias, “The noisy channel coding theorem for erasure channels,” The American Mathematical Monthly, pp. 853–862, 1974.