On the Optimality of Treating Interference as Noise: General Message Sets

Chunhua Geng, Student Member, IEEE, Hua Sun, Student Member, IEEE, and Syed Ali Jafar, Fellow, IEEE

Abstract—In a K-user Gaussian interference channel, it has been shown that if for each user the desired signal strength is no less than the sum of the strengths of the strongest interference from this user and the strongest interference to this user (all values in decibel scale), then treating interference as noise (TIN) is optimal from the perspective of generalized degrees of freedom (GDoF) and achieves the entire channel capacity region to within a constant gap. In this paper, we show that for such TIN-optimal interference channels, even if the message set is expanded to include an independent message from each transmitter to each receiver, operating the new channel as the original interference channel and treating interference as noise is still optimal for the sum capacity up to a constant gap. Furthermore, we extend the result to the sum-GDoF optimality of TIN in the general setting of X channels with arbitrary numbers of transmitters and receivers.

Index Terms—Gaussian networks, generalized degrees of freedom (GDoF), sum capacity, treating interference as noise (TIN), X channels.

I. INTRODUCTION

TREATING interference as noise (TIN) when it is sufficiently weak is an attractive interference management principle for wireless networks in practice due to its simplicity and robustness. Remarkably, TIN is also information-theoretically optimal when the interference is sufficiently weak. This is established in [1]–[6] from an exact capacity perspective, and in [7]–[12] from an approximate capacity perspective. Each approach has its merits – the former identifies relatively narrow regimes where TIN achieves exact capacity, whereas the latter identifies significantly broader regimes where TIN is approximately optimal. Most relevant to this work are the results by Geng et al. in [12] where it is shown that in a general K-user interference channel, if for each user the desired signal strength is no less than the sum of the strengths of the strongest interference from this user and the strongest interference to this user (all values in dB scale), then TIN achieves the entire generalized degrees of freedom (GDoF) region and is optimal for the whole channel capacity region up to a constant gap of no more than \( \log_2(3K) \) bits per user.

In this paper we explore the sum-rate optimality of TIN when the message set is expanded to include an independent message from each transmitter to each receiver, operating as the original interference channel and treating interference as noise at each receiver is still optimal for the sum capacity up to a constant gap (see Theorem 2 in Section III). We also extend the optimality of TIN to the general X channel with arbitrary numbers of transmitters and receivers (see Theorem 3 in Section III). In each case, the achievability argument follows directly from [12] because it is based only on operating the target network as an interference channel and treating interference as noise. The main difficulty lies in deriving tight information theoretical outer bounds. Recall that in [12], for interference channels, the converse is based on reducing the channel to a cyclic network. Each such reduction produces an outer bound and collectively these outer bounds suffice for the GDoF characterization in the interference channel setting. However, this is no longer true once the message set is expanded. While one can similarly obtain outer bounds on the sum rates of subsets of messages by considering all cyclic subnetworks, it is not hard to verify that these bounds do not suffice for our purpose. As illustrated in this work, to obtain desired outer bounds, the genie signal provided to each user should be chosen more judiciously. Notably, to complete the generalization to X channels, especially for the case where the number of receivers is larger than that of transmitters, due to the added difficulty, we also resort to deterministic channel models [16]–[22], in which certain combinatorial structure can be exploited to simplify the proof. More specifically, we use the fact that the sum capacity of Gaussian X channels is upper bounded by that of their carefully chosen deterministic counterparts up to a constant gap, which shares the same...
principle with [20]–[22]. Then by upper bounding the sum capacity of the deterministic channel appropriately, we obtain the desired converse for the original Gaussian channel.

II. PRELIMINARIES

A. Channel Model

Consider the wireless channel with $M$ transmitters and $N$ receivers, which can be described by the following input-output equations,

$$Y_k(t) = \sum_{i=1}^{M} \tilde{h}_{ki} X_i(t) + Z_k(t), \quad \forall k \in [N],$$

where $\tilde{h}_{ki}$ is the complex channel gain value from transmitter $i$ to receiver $k$. $X_i(t)$ and $Z_k(t)$ are the transmitted symbol of transmitter $i$, the received signal of receiver $k$, and the additive circularly symmetric complex Gaussian noise with zero mean and unit variance seen by receiver $k$, respectively, at each time index $t$. All the symbols are complex. Each transmitter $i$ is subject to the power constraint $E[|\tilde{X}_i(t)|^2] \leq P_i$.

Following similar approaches in [7] and [12], we translate the standard channel model (1) into an equivalent normalized form that is more conducive for GDoF studies. We define

$$a_{ki} \triangleq \frac{\log(1 + [|\tilde{h}_{ki}|^2 P_i])}{\log P}, \quad \forall i \in [M], \quad \forall k \in [N],$$

where $P > 1$ is a nominal power value.

Now according to (2), we represent the original channel model (1) in the following form,

$$Y_k(t) = \sum_{i=1}^{M} h_{ki} X_i(t) + Z_k(t) = \sum_{i=1}^{M} \sqrt{P_{ki}} e^{j\theta_{ki}} X_i(t) + Z_k(t), \quad \forall k \in [N],$$

where $X_i(t) = \tilde{X}_i(t)/\sqrt{P_i}$ is the normalized transmit symbol of transmitter $i$, subject to the unit power constraint, i.e., $E[|X_i(t)|^2] \leq 1$. $\sqrt{P_{ki}}$ and $\theta_{ki}$ are the magnitude and the phase, respectively, of the channel between transmitter $i$ and receiver $k$. The exponent $a_{ki}$ is called the channel strength level of the link between transmitter $i$ and receiver $k$. As in [7] and [12], for the GDoF metric, we preserve the ratios $a_{ki}$ as all SNR approaches infinity. In the rest of the paper, we only consider the equivalent channel model in (3).

In the $K$-user interference channel where $M = N = K$, each transmitter intends to send one independent message to its corresponding receiver. Because we wish to prove the negative result that additional messages do not add to the sum-GDoF in a TIN-optimal network, the strongest result corresponds to the case where we include messages from every transmitter to every receiver. Therefore, we will consider the $X$ channel setting. In the $M \times N \times X$ channel, transmitter $i$ has message $W_{ki}$ intended for receiver $k$, and all the messages $W_{ki}$ are independent, $\forall i \in [M], \forall k \in [N]$. The size of the message set $\{W_{ki}\}$ is denoted by $|W_{ki}|$. For codewords spanning a channel uses, the rates $R_{ki} = \log |W_{ki}|/n$ are achievable if the probability of error of all messages can be made arbitrarily small simultaneously by choosing an appropriately large $n$. The channel capacity region $C$ is the closure of the set of all achievable rate tuples. Collecting the channel strength levels and phases in the sets

$$a \triangleq \{a_{ki}\}, \quad \theta \triangleq \{\theta_{ki}\}, \quad \forall i \in [M], \quad \forall k \in [N],$$

the capacity region is denoted as $C(P, a, \theta)$, which is a function of $a$, $\theta$, and $P$. The sum channel capacity is defined as

$$C_{\Sigma, X} \triangleq \max_{C(P, a, \theta)} \sum_{i=1}^{M} \sum_{k=1}^{N} R_{ki}$$

Then the GDoF region of the $X$ channel as represented in (3) is given by

$$D(\alpha, \theta) \triangleq \left\{(d_{11}, d_{12}, \ldots, d_{NM}) : d_{ki} = \lim_{P \to \infty} \frac{R_{ki}}{\log P}, \forall i \in [M], \forall k \in [N], (R_{11}, R_{12}, \ldots, R_{NM}) \in C(P, a, \theta) \right\},$$

and its sum-GDoF value is

$$d_{\Sigma, X} \triangleq \max_{D(\alpha, \theta)} \sum_{i=1}^{M} \sum_{k=1}^{N} d_{ki}$$

B. On the Optimality of TIN for Interference Channels

Let us first review the optimality of TIN for the $K$-user interference channel from the perspective of GDoF.

Theorem 1 [12, Th. 1]: In a $K$-user interference channel, where the channel strength level from transmitter $i$ to receiver $j$ is equal to $a_{ji}$, $\forall i, j \in [K]$, if the following condition is satisfied

$$a_{ii} \geq \max_{j:j \neq i} \{a_{ji}\} + \max_{k:k \neq i} \{a_{ik}\}, \quad \forall i, j \in [K],$$

then power control and treating interference as noise achieves the whole GDoF region. Moreover, the GDoF region is the set of all $K$-tuples $(d_1, d_2, \ldots, d_K)$ satisfying

individual bounds : $0 \leq d_i \leq a_{ii}, \quad \forall i \in [K]$ (9)

cycle bounds : $\sum_{j=1}^{m} d_{ij} \leq \sum_{j=1}^{m} (a_{ij} - a_{ij-1}), \forall i \in [1, 2, \ldots, m], \forall m \in [2, 3, \ldots, K]$ (10)

where $\Pi_K$ is the set of all possible cyclic sequences of all subsets of $[K]$ with cardinality no less than 2, and the modulo-$m$ arithmetic is implicitly used on user indices, e.g., $i_m = i_0$.

To help interpret the results in Theorem 1, consider a 3-user interference channel as an example, where

$$\Pi_3 = \{(1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\}$$
Note that each cyclic sequence in $\Phi_K$ is essentially a cyclically ordered subset of user indices, without repetitions. Then according to Theorem 1, under the condition (8), the GDoF region $(d_1, d_2, d_3)$ is fully characterized by the following inequalities together

$$
0 \leq d_1 \leq a_{11},
0 \leq d_2 \leq a_{22},
0 \leq d_3 \leq a_{33},
d_1 + d_2 \leq (a_{11} + a_{22}) - (a_{12} + a_{21}),
d_1 + d_3 \leq (a_{11} + a_{33}) - (a_{13} + a_{31}),
d_2 + d_3 \leq (a_{22} + a_{33}) - (a_{23} + a_{32}),
d_1 + d_2 + d_3 \leq (a_{11} + a_{22} + a_{33}) - (a_{12} + a_{23} + a_{31}),
d_1 + d_2 + d_3 \leq (a_{11} + a_{22} + a_{33}) - (a_{13} + a_{32} + a_{21})
$$

which is achievable via power control and TIN.

Remark: Theorem 1 claims that in the $K$-user interference channel, if for each user the desired signal strength is no less than the sum of the strengths of the strongest interference from this user and the strongest interference to this user (all values in dB scale), then TIN is GDoF-optimal. Furthermore, it is shown in [12] that under the same condition, TIN achieves the entire channel capacity region to within a gap of no larger than $ \log_2 (3K)$ bits per user. Note that the gap is bounded by a constant for a fixed number of users, i.e., it does not depend on the channel strength parameters $a_{ij}$ and $P$.

### III. Results

The main result of this paper is the following theorem.

**Theorem 2:** In a $K$-user interference channel, where the channel strength level from transmitter $i$ to receiver $j$ is equal to $a_{ij}$, $\forall i, j \in [K]$, when the following condition is satisfied,

$$
a_{ii} \geq \max_{j, k \neq i} [a_{ij} + \max_{k \neq i} a_{ik}], \quad \forall i, j, k \in [K] \quad (11)
$$

then even if the message set is increased to the $X$ channel setting, operating the new channel as the original interference channel and treating interference as noise at each receiver still achieves the sum-GDoF. Furthermore, the same scheme is also optimal for the sum channel capacity up to a constant gap of no more than $K \log_2 [K(K + 1)]$ bits.

The proof of Theorem 2 is presented in Section IV-A.

While the $K$-user interference channel is naturally associated with a $K \times K$ $X$ channel setting, the $X$ channel setting also allows for unequal numbers of transmitters and receivers. A natural question is whether such a generalization of the TIN-optimality result is possible for the $X$ channel with $M \neq N$. The following theorem provides such a generalization.

**Theorem 3:** In an $M \times N$ $X$ channel, where the channel strength level from transmitter $i$ to receiver $j$ is equal to $a_{ij}$ and $\kappa \triangleq \min(M, N)$, if there exist two permutations $\Pi^T$ and $\Pi^R$ for the transmitter and receiver indices, respectively, such that

$$
a_{\Pi^T_i, \Pi^R_j} \geq \max_{j, k \neq i} [a_{\Pi^T_j, \Pi^R_j}], \quad \forall i \in [\kappa], \quad \forall j \in [N], \quad \forall k \in [M], \quad (12)
$$

then according to Theorem 1, under the condition (11), the sum-GDoF value is still

$$
\sum_{i=1}^{\kappa} \sum_{k=1}^{M} d_{ki} = 2.5
$$

which is achieved by power control and TIN.

Next, let us expand the set of messages to the $X$ channel setting, where each transmitter intends to send an independent message to each receiver as shown in Fig. 2. Therefore, there are totally 9 messages in this $X$ channel. Theorem 2 claims that for this $3 \times 3$ $X$ channel, the sum-GDoF value is still

$$
\sum_{i=1}^{3} \sum_{k=1}^{3} d_{ki} = 2.5
$$

which can be achieved by setting $W_{ki} = \phi$ for $i \neq k$ and $\forall i, k \in \{1, 2, 3\}$, sending only $\{W_{11}, W_{22}, W_{33}\}$ through the channel and treating interference as noise at each receiver.

---

Fig. 1. A 3-user interference channel, where the value on each link denotes its channel strength level.

Fig. 2. A $3 \times 3$ $X$ channel, which has the same channel strength values as the 3-user interference channel in Fig. 1.

where $\Pi^T_i$ ($\Pi^R_i$) denotes the $i$-th element in the permutation of transmitters (receivers) $\Pi^T$ ($\Pi^R$), then operating the channel as a $\kappa$-user interference channel, in which each transmitter $\Pi^T_i$ intends to deliver an independent message to receiver $\Pi^R_i$, $i \in [\kappa]$, and treating interference as noise at each receiver is sum-GDoF optimal.\(^3\)

The proof of Theorem 3 is given in Section IV-B.

**Example 1:** First, consider the 3-user interference channel illustrated in Fig. 1, where the value on each link denotes its channel strength level. In this interference channel, transmitter $i$ intends to send an independent message $W_i$ to its desired receiver $i$, $\forall i \in \{1, 2, 3\}$. Hence there are 3 messages in this setting. It’s easy to check that the TIN-optimality condition (11) is satisfied for each user. Then according to Theorem 1, it is not hard to verify that the sum-GDoF value of this interference channel is

$$
d_{\Sigma, 1C} = d_1 + d_2 + d_3 = 2.5
$$

\(^3\)Based on the proof in Section IV-B, it is not hard to verify that the same TIN scheme is also optimal to achieve the sum channel capacity to within a constant gap.
Finally, after adding another transmitter as depicted in Fig. 3, the number of the messages increases to 12 in this $4 \times 3$ X channel. It’s easy to verify that (12) holds. Then according to Theorem 3, for this X channel and its reciprocal channel, the same TIN scheme is still optimal in terms of the sum-GDoF, whose value remains 2.5.

IV. PROOFS

A. Proof for Theorem 2

In the following, we first consider the sum-GDoF of the $K \times K$ X channel. Then we use the insight gained in the GDoF study to derive the constant gap result for the sum channel capacity.

1) Proof for the Sum-GDoF: The proof consists of two steps. First, we show that when the condition (11) is satisfied, for all individual and cycle bounds of a TIN-optimal $K$-user interference channel (see Theorem 1), if each $d_i$ ($\forall i \in \{K\}$) is replaced by $\bar{d}_i = \sum_{j=1}^{K} d_{ij}$, these bounds still hold for its counterpart X channel. Next, based on the first step, we prove that under the condition (11), the $K$-user interference channel and its counterpart X channel have the same sum-GDoF. Therefore, according to Theorem 1, we establish that power control and TIN achieves the sum-GDoF of the $K \times K$ X channel when the condition (11) holds.

Let’s start with the first step. In the following, we first give an example of the $3 \times 3$ X channel to show how to obtain the outer bounds corresponding to the individual and cycle bounds in Theorem 1, where each $d_i$ is replaced by $\bar{d}_i = \sum_{j=1}^{3} d_{ij}$, $\forall i \in \{1, 2, 3\}$. Then we generalize the proof to the $K \times K$ X channel.

Example 2: Consider a 3-user TIN-optimal interference channel. According to Theorem 1, we can obtain the entire GDoF region, which is characterized by certain individual and cycle bounds. To extend the result to the X channel setting, each of these bounds will be extended. To illustrate these key ideas in this example, we consider the following two bounds,

\[ d_3 \leq \alpha_{33} \quad (13) \]
\[ d_1 + d_2 \leq (\alpha_{11} + \alpha_{22}) - (\alpha_{12} + \alpha_{21}) \quad (14) \]

and intend to prove that in the counterpart $3 \times 3$ X channel, if we replace each $d_i$ by $\bar{d}_i = \sum_{j=1}^{3} d_{ij}$, $\forall i \in \{1, 2, 3\}$, the above two bounds still hold, i.e.,

\[ \hat{d}_3 = d_{31} + d_{32} + d_{33} \leq \alpha_{33} \quad (15) \]
\[ \hat{d}_1 + \hat{d}_2 = d_{11} + d_{12} + d_{13} + d_{21} + d_{22} + d_{23} \leq (\alpha_{11} + \alpha_{22}) - (\alpha_{12} + \alpha_{21}) \quad (16) \]

All the remaining bounds can be extended to the X channel similarly.

To prove (15), we just need to consider the multiple-access channel (MAC) consisting of all the transmitters and the receiver 3, then we have

\[ R_{31} + R_{32} + R_{33} \leq 2 \log_2(1 + P_{\alpha_{31}} + P_{\alpha_{32}} + P_{\alpha_{33}}) \quad (17) \]

Because (11) is satisfied, i.e., $\alpha_{33} \geq \alpha_{32}$ and $\alpha_{33} \geq \alpha_{31}$, in the GDoF sense we have

\[ \hat{d}_3 = d_{31} + d_{32} + d_{33} \leq \alpha_{33} \quad (18) \]

To prove (16), consider the subnetwork consisting of all the transmitters and the receivers 1 and 2 as illustrated in Fig. 4, where we have eliminated the third receiver and its desired messages $W_{31}, W_{32}, W_{33}$. This cannot hurt the rates of the remaining messages, so the outer bound arguments remain valid. Define

\[ S_1(t) = h_{21} X_1(t) + Z_2(t) \quad (19) \]
\[ S_2(t) = h_{12} X_2(t) + Z_1(t) \quad (20) \]

For the receiver 1, we provide $S_1^n, W_{21}$ and $W_{23}$ through a genie. From Fano’s inequality, we have

\[ n(R_{11} + R_{12} + R_{13} - \epsilon) \]
\[ \leq I(W_{11}, W_{12}, W_{13}; Y^n_{1}, S^n_{1}, W_{21}, W_{23}) \quad (21) \]
\[ = I(W_{11}, W_{12}, W_{13}; Y^n_{1}, S^n_{1}|W_{21}, W_{23}) \quad (22) \]
\[ = I(W_{11}, W_{12}, W_{13}; S^n_{1}|W_{21}, W_{23}) + I(W_{11}, W_{12}, W_{13}; Y^n_{1}|W_{21}, W_{23}) \quad (23) \]
\[ = h(S^n_{1}|W_{21}, W_{23}) - h(S^n_{1}|W_{21}, W_{23}, W_{11}, W_{12}, W_{13}) + h(Y^n_{1}|S^n_{1}, W_{21}, W_{23}) \quad (24) \]
\[ \leq h(S^n_{1}|W_{21}, W_{23}) - h(S^n_{1}|W_{21}, W_{23}, W_{11}, W_{12}, W_{13}, X^n_{2}) + h(Y^n_{1}|S^n_{1}, W_{21}, W_{23}) \quad (25) \]
\[ \leq h(S^n_{1}|W_{21}) - h(Z^n_{2}) + h(Y^n_{1}|S^n_{1}) - h(S^n_{2}|W_{12}) \quad (26) \]

where (22) follows because all the messages are independent. (25) holds since adding conditioning does not increase
Due to the condition (11), in the GDoF sense we obtain
\[
\begin{align*}
 n (R_{21} + R_{22} + R_{23} - \epsilon) \\
\leq h(S'_{21}^n | W_{12}) - h(Z'_{1}^n) + h(Y'_{2}^n | S'_{22}^n) - h(S'_{23}^n | W_{21})
\end{align*}
\] (27)

Thus the sum rate is bounded as follows.
\[
\begin{align*}
 n(\sum_{i=1}^{2} \sum_{j=1}^{3} R_{ij} - 2\epsilon) \\
\leq h(Y'_{2}^n | S'_{1i}^n) + h(Y'_{2}^n | S'_{22}^n) - h(Z'_{1}^n) - h(Z'_{2}^n)
\end{align*}
\] (28)

\[
\begin{align*}
\leq \sum_{i=1}^{n} \left[ h(Y_{1}(t) | S_{1i}(t)) + h(Y_{2}(t) | S_{2i}(t)) - h(Z_{1}(t)) \\
- h(Z_{2}(t)) \right]
\end{align*}
\] (29)

where the second inequality follows from the chain rule and the fact that dropping conditioning does not reduce entropy. Finally, because the circularly symmetric complex Gaussian distribution maximizes conditional differential entropy for a given covariance constraint, we obtain
\[
\begin{align*}
2 \sum_{i=1}^{2} \sum_{j=1}^{3} R_{ij} - 2\epsilon \\
\leq \log_{2} \left( 1 + P^{a_{13}} + P^{a_{12}} + \frac{P^{a_{11}}}{1 + P^{a_{21}}} \right) \\
+ \log_{2} \left( 1 + P^{a_{23}} + P^{a_{22}} + \frac{P^{a_{22}}}{1 + P^{a_{12}}} \right)
\end{align*}
\] (30)

Due to the condition (11), in the GDoF sense we obtain
\[
\begin{align*}
\hat{d}_{1} + \hat{d}_{2} \leq (a_{11} + a_{22}) - (a_{12} + a_{21})
\end{align*}
\] (31)

which is the desired extension, (16), to the X channel setting of the original bound, (14), for the interference channel.

Now let us consider the proof for the general $K \times K$ X channel. For the individual bounds in the $K$-user interference channel
\[
d_i \leq a_{ii}, \quad \forall i \in [K],
\] (32)

in its counterpart X channel, the corresponding bound comes from the MAC consisting of all the transmitters and the receiver i,
\[
\sum_{j=1}^{K} R_{ij} \leq \log_{2} (1 + \sum_{j=1}^{K} P^{a_{ij}})
\] (33)

According to (11), in the GDoF sense we have
\[
\begin{align*}
\hat{d}_{i} = \sum_{j=1}^{K} d_{ij} \leq a_{ii}
\end{align*}
\] (34)

For any cycle bound in the interference channel
\[
\begin{align*}
\sum_{j=1}^{m} d_{ij} \leq \sum_{j=1}^{m} (a_{ij} - a_{ij-1}), \\
\forall (i_1, i_2, \ldots, i_m) \in \Pi_K, \quad \forall m \in \{2, 3, \ldots, K\},
\end{align*}
\] (35)

consider the subnetwork consisting of all the transmitters and the receivers $\{i_1, i_2, \ldots, i_m\}$ as shown in Fig. 5. Eliminate all other receivers and their desired messages, which cannot hurt the rates of the remaining messages. For such a $K \times m$ X channel, define $\mathcal{W} = \{W_{ij_{1k}}\}$, $\mathcal{W}^{*}_{ij_{1k}} = \{W_{ij_{1k}}, W_{ij_{2k}}, \ldots, W_{ij_{jk}}\}$, $\mathcal{W}^{\perp}_{ij_{1k}} = \{W_{ij_{1k}}, W_{ij_{2k}}, \ldots, W_{ij_{jk}}\}$, and $\mathcal{W}^{\perp}_{ij_{k}} = \mathcal{W} \setminus \mathcal{W}_{S}$, where $\forall j \in [m]$, $\forall k \in [K]$, and $S$ is any subset of message indices. In words, the sets $\mathcal{W}$, $\mathcal{W}^{*}_{ij_{1k}}$, and $\mathcal{W}^{\perp}_{ij_{1k}}$ represent all the messages delivered in this $K \times m$ X channel, all the messages intended to receiver $ij_k$, and all the messages coming from transmitter $ik$, respectively, and $\mathcal{W}^{\perp}_{ij_{k}}$ is the complement of $\mathcal{W}^{\perp}_{ij_{k}}$ in $\mathcal{W}$. For instance, when $j, k \in \{1, 2\}$ and $S = \{i_{11}, i_{12}\}$, then $\mathcal{W}^{\perp}_{ij_{1k}} = \{W_{i_{11}1}, W_{i_{12}1}\}$ and $\mathcal{W}^{\perp}_{ij_{k}} = \{W_{i_{11}1}, W_{i_{12}1}\}$.

Modulo-$m$ arithmetic is used on the receiver indices, e.g., $i_0 = i_m$. Lastly, to complete the setup, define
\[
S_{ij}(t) = h_{ij_1,ij} X_{ij}(t) + Z_{ij_1,ij}(t), \quad \forall j \in [m]
\] (36)

For the receiver $i_1$, we provide $S_{i_1}^{\text{err}}$, $\mathcal{W}^{\perp}_{i_2} \setminus \mathcal{W}^{\perp}_{i_1}$ through a genie. From Fano’s inequality, we have
\[
\begin{align*}
n(\sum_{k=1}^{K} R_{ij_{1k}} - \epsilon) \\
\leq I(\mathcal{W}^{\perp}_{i_2}; Y_{i_1} | S_{i_1}^{\text{err}}) - h(S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2})
\end{align*}
\] (37)

\[
\begin{align*}
= I(\mathcal{W}^{\perp}_{i_2}; Y_{i_1} | S_{i_2}^{\text{err}}) - h((\mathcal{W}^{\perp}_{i_1} \setminus \mathcal{W}^{\perp}_{i_2}) | \mathcal{W}^{\perp}_{i_2})
\end{align*}
\] (38)

\[
\begin{align*}
= I(\mathcal{W}^{\perp}_{i_2}; S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2}) + I(\mathcal{W}^{\perp}_{i_2}; S_{i_2}^{\text{err}} | \mathcal{W}^{\perp}_{i_2} | \mathcal{W}^{\perp}_{i_2})
\end{align*}
\] (39)

\[
\begin{align*}
= h(S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2} | \mathcal{W}^{\perp}_{i_2}) - h(S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2})
\end{align*}
\] (40)

\[
\begin{align*}
\leq h(S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2} | \mathcal{W}^{\perp}_{i_2}) - h(Z_{i_1}^{\text{err}}) + h(S_{i_1}^{\text{err}} | \mathcal{W}^{\perp}_{i_2})
\end{align*}
\] (41)

where (38) follows because all the messages are independent, and in (41) we use the fact that dropping conditioning does not reduce entropy.

Similarly, for other receivers $ij$, $\forall j \in \{2, 3, \ldots, m - 1\}$, by providing $S_{ij}^{\text{err}}$, $\mathcal{W}^{\perp}_{ij_{j+1}} \setminus \mathcal{W}^{\perp}_{ij_{j+1}}$ through a genie we have
\[
\begin{align*}
n(\sum_{k=1}^{K} R_{ij_{jk}} - \epsilon) \leq h(S_{ij}^{\text{err}} | \mathcal{W}^{\perp}_{ij_{j+1}} | \mathcal{W}_{ij_{j+1}}) - h(S_{ij}^{\text{err}})
\end{align*}
\] (42)
Finally, for the receiver $i_m$, we provide $S_{m}^{i},\mathcal{W}_{i_{1}}^{\mathcal{X}_{i_{1}}} \mathcal{W}_{i_{m}}^{n}$ through a genie and obtain

$$n\left(\sum_{k=1}^{n}R_{i_{m}(k)} - \epsilon\right) \leq h(S_{m}^{i} | \mathcal{W}_{i_{1}}^{m} | \mathcal{W}_{i_{m}}^{n}) - h(Z_{i_{m}}^{n})$$

$$+ h(Y_{i_{m}}^{n} | S_{m}^{i}) - h(S_{i_{1}}^{n} | \mathcal{W}_{i_{1}}^{m})$$

(43)

Then taking the sum of $n(\sum_{k=1}^{n}R_{i_{m}(k)} - \epsilon)$ for all $j \in [m]$, we have

$$n\left(\sum_{j=1}^{m} \sum_{k=1}^{K} R_{i_{j}(k)} - m\epsilon\right) \leq \sum_{j=1}^{m} \left[h(Y_{j}^{n} | S_{j}^{n}) - h(Z_{j}^{n})\right]$$

$$\leq \sum_{j=1}^{m} \sum_{k=1}^{K} \left[h(Y_{j}(t) | S_{j}(t)) - h(Z_{j}(t))\right]$$

(44)

(45)

where (45) follows the chain rule and the fact that dropping conditioning does not reduce entropy. Again, using the fact that the circularly symmetric complex Gaussian distribution maximizes conditional differential entropy for a given covariance constraint and the condition (11), we can obtain the following desired outer bound in the GDoF sense, through the same set of manipulations as in Example 2,

$$\sum_{j=1}^{m} d_{ij} = \sum_{j=1}^{m} \sum_{k=1}^{K} d_{i_{j}(k)} \leq \sum_{j=1}^{m} (a_{ij} - a_{i_{j-1}i_{j}})$$

(46)

Now we can proceed to the second step to prove that under the condition (11), the $K$-user interference channel and its counterpart $K \times K$ $X$ channel have the same sum-GDoF. According to Theorem 1, for the $K$-user interference channel, under the condition (11), to obtain its sum-GDoF $d_{X,IC}$, we need to solve the following linear programming (LP) problem

$$\max \sum_{i=1}^{K} d_{i}$$

s.t. $0 \leq d_{i} \leq a_{ii}, \forall i \in [K]$ (47)

$$\sum_{j=1}^{m} d_{ij} \leq \sum_{j=1}^{m} (a_{ij} - a_{i_{j-1}i_{j}}), \forall (i_{1}, i_{2}, \ldots, i_{m}) \in \Pi_{K}, \forall m \in \{2, 3, \ldots, K\}$ (48)

To get the sum-GDoF of its counterpart $X$ channel $d_{X,X}$, we consider a similar LP problem. Note for this LP problem with the objective function $\sum_{i=1}^{K} d_{i}$, from the first step, we know that at least it needs to follow two similar constraints to (48) and (49), in which each $d_{i}$ is just replaced by $d_{i}$. Thus we have $d_{X,IC} \geq d_{X,X}$. Obviously, in any case, the sum-GDoF of the $K$-user interference channel must be less than or equal to that of its counterpart $X$ channel, i.e., $d_{X,IC} \leq d_{X,X}$. Therefore, under the condition (11), we have established that the $K$-user interference channel and its counterpart $X$ channel have the same sum-GDoF.

2) Proof for the Constant Gap of Sum Capacity: Based on the insight gained in the above GDoF study, for the TIN-optimal $K \times K$ $X$ channel, we intend to characterize the sum channel capacity to within a constant gap of no more than $K \log_{2}(K + 1)$ bits. To this end, first recall the achievability proof in [12]. By operating the $K \times K$ $X$ channel as an interference channel, in which each transmitter $i$ sends one independent message $W_{i}$ to its corresponding receiver $i$ ($\forall i \in [K]$), power control and TIN can achieve the following rate tuples $(R_{1,TIN}, R_{2,TIN}, \ldots, R_{K,TIN})$ satisfying

$$0 \leq R_{i,TIN} \leq \log_{2}\left(1 + \frac{1}{K}\right), \forall i \in [K]$$

(50)

$$0 \leq R_{i,TIN} \leq \log_{2}\left(1 + \frac{1}{K}\right), \forall i \in [K]$$

(51)

for all cyclic sequences $(i_{1}, i_{2}, \ldots, i_{m}) \in \Pi_{K}$, $\forall m \in \{2, 3, \ldots, K\}$.

Next, consider the converse. Define $\hat{R}_{i} = \sum_{j=1}^{K} R_{ij}$. Start with the individual bounds,

$$\hat{R}_{i} \leq \log_{2}\left(1 + \sum_{j=1}^{K} P a_{ij}\right)$$

(52)

$$\leq \log_{2}\left(K + 1\right) P a_{ij}$$

(53)

$$= a_{ii} \log_{2} P + \log_{2}(K + 1)$$

(54)

Then for the cycle bounds, from (45), it is easy to obtain

$$\sum_{j=1}^{m} \hat{R}_{ij} \leq \sum_{j=1}^{m} \left[\left(a_{ij} - a_{i_{j-1}i_{j}}\right) \log_{2} P + \log_{2}(K + 1)\right]$$

(55)

(56)

for all cyclic sequences $(i_{1}, i_{2}, \ldots, i_{m}) \in \Pi_{K}$, $\forall m \in \{2, 3, \ldots, K\}$.

Compare (50) with (54). Denote the difference between the achievable rate in (50) and the outer bound in (54) as $\delta_{\hat{R}_{i}}$. Consider the following two cases,

- $a_{ii} \log_{2} P + \log_{2}(K + 1) > 0$: In this case, we have

$$\delta_{\hat{R}_{i}} = \left[a_{ii} \log_{2} P + \log_{2}(K + 1)\right]$$

$$- \left[0, a_{ii} \log_{2} P + \log_{2}\left(\frac{1}{K}\right)\right]$$

(57)

$$= a_{ii} \log_{2} P + \log_{2}(K + 1) - a_{ii} \log_{2} P + \log_{2}\left(\frac{1}{K}\right)$$

(58)

$$= \log_{2}(K + 1)$$

- $a_{ii} \log_{2} P + \log_{2}(K + 1) \leq 0$: In this case, we have

$$\delta_{\hat{R}_{i}} = \left[a_{ii} \log_{2} P + \log_{2}(K + 1)\right]$$

$$- \left[0, a_{ii} \log_{2} P + \log_{2}\left(\frac{1}{K}\right)\right]$$

(59)

$$= a_{ii} \log_{2} P + \log_{2}(K + 1) - a_{ii} \log_{2} P + \log_{2}\left(\frac{1}{K}\right)$$

(60)

$$= \log_{2}(K + 1)$$
In both cases, we obtain
\[ \delta_{\hat{R}_i} \leq \log_2[K(K + 1)], \quad \forall i \in [K] \] (57)

Similarly, denote the difference between the achievable rate in (51) and the outer bound in (56) as \( \delta_{\sum_{j=1}^{m} \hat{R}_j} \). Comparing (51) with (56), we always have
\[ \delta_{\sum_{j=1}^{m} \hat{R}_j} \leq m \log_2[K(K + 1)] \] (58)

for all cyclic sequences \( (i_1, i_2, \ldots, i_m) \in \Pi_K, \forall m \in \{2, 3, \ldots, K\} \).

According to (57) and (58), we can characterize the sum channel capacity to within a constant gap of no more than \( K \log_2[K(K + 1)] \) bits, which is only dependent on the number of users \( K \).

B. Proof for Theorem 3

It is easy to verify that when \( M \geq N \), by defining \( \hat{d}_i = \sum_{j=1}^{M} d_{ij} \) (\( \forall i \in [N]\)) and following the same argument as in the proof of Theorem 2, we can complete the proof. Therefore, hereafter we only consider the case where \( K = M < N \).

Without loss of generality, we assume that in the two permutations \( \Pi^I \) and \( \Pi^R \) which satisfy the condition (12), \( \Pi^I_i = i, \forall i \in [M] \) and \( \Pi^R_j = j, \forall j \in [N] \), i.e., the following condition is satisfied,
\[ a_{ii} \geq \max_{j: j \neq i} \{ a_{jj} \} + \max_{k: k \neq i} \{ a_{ik} \}, \quad \forall i, k \in [M], \forall j \in [N] \] (59)

In this case, similar to the proof of Theorem 2, the key step is to show that when (59) holds, for each individual bound and cycle bound in the \( M \)-user interference channel consisting of transmitters \( i \in [M] \) and receivers \( j \in [M] \), if each \( d_{ij} \) (\( \forall i \in [M]\)) is replaced by \( d_{ij} = \sum_{j=1}^{N} d_{j|i} \), the resulting bounds hold in the \( M \times N \) channel. Then based on the same argument of Theorem 2, we can prove the optimality of TIN for the sum-GDoF of the \( M \times N \) channel where \( M < N \).

For the individual bounds, consider the degraded broadcast channel (BC) comprised of the transmitter \( i \) (\( \forall i \in [M] \)) and all the receivers, eliminating all other transmitters and their messages. Since (59) is satisfied, the receiver \( i \) is the strongest receiver, and can decode all the messages from the transmitter \( i \). Thus in the GDoF sense we have
\[ \hat{d}_i = \sum_{j=1}^{N} d_{j|i} \leq a_{ii}, \quad \forall i \in [M] \] (60)

Now the only task left is to prove that in the \( X \) channel setting, by replacing \( d_{i} \) with \( \hat{d}_i \), all the cycle bounds still hold. Before exploring the proof details, let’s see an intuitive sketch of proof first for a 2 x 4 X channel as illustrated in Fig. 6. For this 2 x 4 channel, when (59) is satisfied, we intend to prove
\[ \hat{d}_1 + \hat{d}_2 \leq (a_{11} + a_{22}) - (a_{12} + a_{21}) \]

\[ \text{For the receiver 1, we provide} \ S_1^1 \text{and} \ W_1^1 \text{through a genie. From Fano’s inequality, we obtain} \]
\[ n(R_{11} + R_{12} - \epsilon) \]
\[ \leq I(W_{11}, W_{12}; Y_{11}^1, S_1^1, W_1^1) \] (61)
\[ = I(W_{11}, W_{12}; Y_{11}^1, S_1^1 | W_2^1) \] (62)
\[ = h(Y_{11}^1, S_1^1 | W_2^1) - h(Y_{11}^1, S_1^1 | W_1^1, W_{11}, W_{12}) \] (63)
\[ = h(S_1^1 | W_2^1) + h(Y_{11}^1 | S_1^1, W_2^1) - h(Y_{11}^1 | W_1^1, W_{11}) \] (64)
\[ \leq h(S_1^1 | W_1^1) + h(Y_{11}^1 | S_1^1) - h(Y_{11}^1 | W_1^1, W_{11}) \] (65)
\[ = h(S_1^1 | W_1^1) + h(Y_{11}^1 | S_1^1) - h(Y_{11}^1 | W_1^1, W_{11}) \] (66)
\[ = h(S_1^1 | W_1^1) + h(Y_{11}^1 | S_1^1) - h(Y_{11}^1 | W_1^1, W_{11}, W_{12}, W_{31}, W_{41}) \] (67)
\[ \leq h(S_1^1 | W_1^1) + h(Y_{11}^1 | S_1^1) - h(Y_{11}^1 | W_1^1, W_{12}, W_{31}, W_{41}) \] (68)

where (62) holds since \( W_1^1 \) is independent of \( W_1, W_{12}, W_{31}, W_{41} \), and (66) follows that dropping conditioning (in the second term) does not reduce entropy and adding conditioning (in the last term) does not increase entropy, (66) holds since we can reconstruct \( Y_{11}^1 \) from \( W_1, W_{12}, W_{31} \) and \( W_{41} \). In (68) \( Y_{11}^1, W_{11}^1, W_{12} \) denotes the complement of \( (W_1, W_{12}, W_{31}, W_{41}) \). The last inequality in (68) is the key step of the proof. Intuitively, it is due to the fact that out of the \( a_{11} \) log \( P \) bit levels of \( Y_1 \) that are above the noise floor, \( S_2 \) is contained in the lowest \( a_{12} \) log \( P \) bit levels of \( Y_1 \), whereas only the top \( a_{31} \) log \( P \) bit levels are seen by the receiver 3. Since \( a_{11} \geq a_{12} + a_{31} \), these bit levels do not overlap, i.e., they can be recovered from \( Y_1 \) within a bounded entropy gap. As shown later, this argument becomes evident in a deterministic approach.
Next, consider the degraded BC comprised of the transmitter 1 and the receivers 3 and 4. Since \( a_{31} \geq a_{41} \), we have

\[
n(R_{31} + R_{41} - \epsilon) \leq I(W_{31}, W_{41}; Y_{3}^{c}|V_{31,41}^{c}) = h(Y_{3}^{c}|V_{31,41}^{c}) - h(Z_{3}^{c})
\]

Adding (68) and (70), we obtain

\[
n(R_{11} + R_{12} + R_{31} + R_{41} - 2\epsilon) \leq h(Y_{1}^{c}|S_{11}^{c}) + h(Y_{1}^{c}|S_{12}^{c}) - h(S_{11}^{c}|W_{12}) - n \log(P)
\]

Similarly, we have

\[
n(R_{21} + R_{22} + R_{32} + R_{42} - 2\epsilon) \leq h(Y_{2}^{c}|S_{21}^{c}) + h(Y_{2}^{c}|S_{22}^{c}) - h(S_{21}^{c}|W_{21}) - n \log(P)
\]

Finally, through adding (71) and (72) together and some other manipulations, we can obtain the desired outer bound,

\[
n(R_{32} - 4\epsilon) \leq h(Y_{3}^{c}|S_{32}^{c}) + h(Y_{3}^{c}|S_{31}^{c}) - n \log(P)
\]

The input-output relationship can be written as the equation (73) given at the top of this page, where \( E[|X(t)|^2] \leq 1 \) and \( Z(t) \sim \mathcal{CN}(0, 1) \). To facilitate the following deterministic approach, by scaling the output, we may set

\[
E[|X(t)|^2] \leq 2, \quad Z(t) \sim \mathcal{CN}(0, 2),
\]

which does not affect the channel capacity of the Gaussian model (73).

In this paper, we consider the deterministic model (74) in top of this page, where \([x] \) is the truncation function which maps \( x \) to its integer part, \( m_{ik}^{R} \equiv \lfloor \log_2 |h_{ik}^{R}| \rfloor \), \( m_{ik}^{I} \equiv \lfloor \log_2 |h_{ik}^{I}| \rfloor \), the real and imaginary parts of the input signal \( X_{i}(t) = \tilde{X}_{i}^{R}(t) + j \tilde{X}_{i}^{I}(t) \) both satisfy the unit peak power constraint, and \( \tilde{X}_{i}^{R,b}(t) (X_{i,b}^{I}(t)) \) is the \( b \)-th bit in the fractional part of \( [\tilde{X}_{i}^{R}(t)] (|X_{i,b}^{I}(t)|) \) in the binary expansion. In the following, to distinguish the deterministic model in (74) from others, we call it the truncated binary-expansion deterministic model. The following lemma shows that the sum capacity of the Gaussian X channel in (73) is upper bounded by that of the truncated binary-expansion deterministic model in (74) up to a constant gap.

**Lemma 1:** The sum capacity of the complex Gaussian \( M \times N \times 1 \) channel is upper bounded by the sum capacity of its corresponding truncated binary-expansion deterministic model up to a constant gap.

The proof for the above lemma follows [20] and is relegated to Appendix A for completeness.

Now define \( m_{ij} \equiv \lfloor \frac{1}{2} \log_2 P_{m_{ij}} \rfloor \). Since \( P > 1 \) and \( a_{ii} \geq a_{ij} + a_{ki}, \forall i \notin \{j, k\} \), we have

\[
\frac{a_{ij}}{2} \log_2 P \geq \lfloor \frac{a_{ij} + a_{ki}}{2} \log_2 P \rfloor \]

\[
\Rightarrow \lfloor \frac{a_{ij}}{2} \log_2 P \rfloor \geq \frac{a_{ij}}{2} \log_2 P - \frac{a_{ki}}{2} \log_2 P
\]

\[
\Rightarrow m_{ij} \geq m_{ij} + m_{ki}, \forall i, j, k, \quad i \notin \{j, k\}
\]

\[
5 \text{ We can write the real-valued signal } |\tilde{X}_{i}^{R}| (|\tilde{X}_{i}^{I}|) \leq 1 \text{ in terms of its binary expansion as}
\]

\[
|\tilde{X}_{i}^{R}| = \sum_{b=1}^{m_{ik}^{R}} \tilde{X}_{i,b}^{R} \gamma^{b}, |\tilde{X}_{i}^{I}| = \sum_{b=1}^{m_{ik}^{I}} \tilde{X}_{i,b}^{I} \gamma^{b}
\]
In order to convey the key ingredients of the proof more clearly, next we give an example for the real Gaussian $2 \times 4$ X channel, and then generalize the proof to the complex Gaussian $M \times N$ ($M < N$) X channel.

Example 3: Consider the real Gaussian X channel with 2 transmitters and 4 receivers, where (59) is satisfied. In this example we still assume $a_{31} \geq a_{41}$ and $a_{42} \geq a_{32}$. As previously mentioned, the proof for all the other cases follows similarly. Also define the message set $\mathcal{W} \triangleq \{W_k\}$, $\forall i \in \{1, 2\}$, $\forall k \in \{1, 2, 3, 4\}$.

Recall that for this $2 \times 4$ X channel, we intend to prove the following cycle bound

$$d_1 + d_2 \leq \frac{1}{2}[(a_{11} + a_{22}) - (a_{12} + a_{21})]$$

where the factor $\frac{1}{2}$ is due to the fact that the Gaussian X channel is real-valued.

We start with the corresponding truncated binary-expansion deterministic model. Define

$$\hat{S}_1(t) = \left[ \text{sign}(\tilde{X}_1(t))h_{21} \sum_{b=1}^{m_{21}} \tilde{X}_{1,b}(t)2^{-b} \right]$$

$$\hat{S}_2(t) = \left[ \text{sign}(\tilde{X}_2(t))h_{12} \sum_{b=1}^{m_{12}} \tilde{X}_{2,b}(t)2^{-b} \right].$$

(78)

(79)

Also define

$$\tilde{X}_{31,5}(t) = \text{sign}(\tilde{X}_1(t)) \sum_{b=1}^{m_{31}} \tilde{X}_{1,b}(t)2^{-b}$$

(80)

Thus the output of the receiver 1 can be written as

$$\hat{Y}_1(t) = \left[ \text{sign}(\tilde{X}_1(t))h_{11} \sum_{b=1}^{m_{11}} \tilde{X}_{1,b}(t)2^{-b} \right]$$

$$+ \left[ \text{sign}(\tilde{X}_2(t))h_{12} \sum_{b=1}^{m_{12}} \tilde{X}_{2,b}(t)2^{-b} \right]$$

(81)

$$= \left[ \text{sign}(\tilde{X}_1(t))h_{11} \sum_{b=1}^{m_{11}} \tilde{X}_{1,b}(t)2^{-b} \right] + \hat{C}_1(t)$$

$$+ \left[ \text{sign}(\tilde{X}_1(t))h_{11} \sum_{b=m_{31}+1}^{m_{11}} \tilde{X}_{1,b}(t)2^{-b} \right] + \hat{S}_2(t)$$

(82)

$$= \left[ h_{11} \tilde{X}_{31,5}(t) \right] + \hat{C}_1(t)$$

$$+ \left[ \text{sign}(\tilde{X}_1(t))h_{11} \sum_{b=m_{31}+1}^{m_{11}} \tilde{X}_{1,b}(t)2^{-b} \right] + \hat{S}_2(t)$$

$$\hat{Y}_{1,t}(t)$$

(83)

where $\hat{C}_1(t)$ may take a value from $\{-1, 0, 1\}$.

For the receiver 1, we have

$$n(R_{11} + R_{12} - \epsilon)$$

$$\leq I(W_{11}, W_{12}; \hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21})$$

$$= H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21})$$

$$- H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{11}, W_{12})$$

$$= H(\hat{S}_1^n | W_{21}) + H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n | \hat{S}_1^n, W_{21})$$

$$- H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n | W_{11}, W_{12})$$

$$\leq H(\hat{S}_1^n | W_{21}) + H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21})$$

$$- H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n | W_{11}, W_{12})$$

$$= H(\hat{S}_1^n | W_{21}) + H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{11}, W_{12})$$

(84)

(85)

(86)

(87)

where (87) follows that dropping conditioning does not reduce entropy. Now consider the last term in (87).

$$H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n | W_{21}, W_{11}, W_{12})$$

$$= H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n | W_{21}, W_{11}, W_{12})$$

$$+ H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21}, W_{11}, W_{12})$$

(88)

(89)

(90)

(91)

where (89) holds since the function $f: \tilde{X}_{31,5} \rightarrow \hat{Y}_{1,u}$ is bijective when $m_{11} \geq m_{31}$, and (91) follows that conditioning on the messages $\hat{Y}_{31,41}^c$, the function $f: \tilde{X}_{31,5} \rightarrow \hat{Y}_3$ is bijective. Note that the equations (88)-(91) correspond to the key step (68) in the intuitive proof given before.

Plugging (91) into (87), we have

$$n(R_{11} + R_{12} - \epsilon) \leq H(\hat{S}_1^n | W_{21}) + H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21})$$

$$- H(\hat{S}_2^n | W_{12}) - H(\hat{Y}_3^n | \hat{Y}_{31,41}^c)$$

(92)

Then, consider the degraded BC comprised of the transmitter 1 and the receivers 3 and 4. Since $m_{31} \geq m_{41}$, we have

$$n(R_{31} + R_{41} - \epsilon) \leq I(W_{31}, W_{41}; \hat{Y}_{3}^n | \hat{Y}_{31,41}^c)$$

$$= H(\hat{Y}_3^n | \hat{Y}_{31,41}^c)$$

(93)

(94)

Combining (92) and (94), we obtain

$$n(R_{11} + R_{12} + R_{31} + R_{41} - \epsilon)$$

$$\leq H(\hat{S}_1^n | W_{21}) + H(\hat{Y}_{1,u}^n, \hat{Y}_{1,l}^n, \hat{C}_1^n, \hat{S}_1^n | W_{21}) - H(\hat{S}_2^n | W_{12})$$

(95)

Similarly, by considering the receiver 2 and the degraded BC comprised of the transmitter 2 and the receivers 3 and 4, we obtain

$$n(R_{21} + R_{22} + R_{32} + R_{42} - \epsilon)$$

$$\leq H(\hat{S}_2^n | W_{12}) + H(\hat{Y}_{2,u}^n, \hat{Y}_{2,l}^n, \hat{C}_2^n, \hat{S}_2^n | W_{12}) - H(\hat{S}_1^n | W_{21})$$

(96)
Adding (95) and (96), the sum capacity of this deterministic $2 \times 4 X$ channel is upper bounded by

$$n(R_{Σ,D} - 4ε)$$ \leq \sum_{i=1}^{n} \left[ H(\hat{Y}_{1,a}(t)|\hat{S}_1(t)) + H(\hat{Y}_{1,b}(t)|\hat{S}_2(t)) + H(\hat{C}(t)) \right]$$

(97)

$$\leq \sum_{i=1}^{n} \left[ H(\hat{Y}_{1,a}(t)|\hat{S}_1(t)) + H(\hat{Y}_{1,b}(t)|\hat{S}_2(t)) + H(\hat{C}(t)) \right]$$

(98)

where the last inequality follows from the chain rule and the fact that dropping conditioning does not reduce entropy.

Next, for the term $H(\hat{Y}_{1,a}(t)|\hat{S}_1(t)) + H(\hat{Y}_{1,b}(t)|\hat{S}_2(t))$, we consider two cases,

- $m_{21} \geq m_{31}$:

$$H(\hat{Y}_{1,a}(t)|\hat{S}_1(t)) + H(\hat{Y}_{1,b}(t)|\hat{S}_2(t)) \leq 0 + (m_{11} - m_{21}) + \text{constant}$$

(99)

$$= (m_{11} - m_{21}) + \text{constant}$$

(100)

where (99) follows that conditioning on $\hat{S}_1$, out of the received signal $\hat{Y}_{1,b}$, both the signals from transmitter 1 and 2 have at most $m_{11} - m_{21}$ bit levels, and the signs of the signals and carry-overs due to the sum of two such signals can only induce a loss of constant bits.

- $m_{21} < m_{31}$: Similarly, we have

$$H(\hat{Y}_{1,a}(t)|\hat{S}_1(t)) + H(\hat{Y}_{1,b}(t)|\hat{S}_2(t)) \leq (m_{22} - m_{12}) + \text{constant}$$

(101)

Due to symmetry, we always have

$$H(\hat{Y}_{2,a}(t)|\hat{S}_2(t)) + H(\hat{Y}_{2,b}(t)|\hat{S}_2(t)) \leq (m_{22} - m_{12}) + \text{constant}$$

(102)

Therefore,

$$n(R_{Σ,D} - 4ε)$$ \leq \sum_{i=1}^{n} \left[ (m_{11} - m_{21}) + (m_{22} - m_{12}) + \text{constant} \right]$$

(103)

According to Lemma 1, for the sum capacity of the original Gaussian $X$ channel $R_{Σ,G}$, we have

$$R_{Σ,G} \leq R_{Σ,D} + \text{constant}$$

(104)

$$\leq (m_{11} - m_{21}) + (m_{22} - m_{12}) + \text{constant}$$

(105)

$$\leq \frac{1}{2}((a_{11} - a_{21}) + (a_{22} - a_{12})) \log_2 P + \text{constant}$$

(106)

Finally, we obtain the desired GDoF cycle bound,

$$\tilde{d}_1 + \tilde{d}_2 \leq \frac{1}{2}((a_{11} - a_{21}) + (a_{22} - a_{12}))$$

(107)

Equipped with the bounding techniques in the above example, we can extend the proof to the complex Gaussian $M \times N (M < N) X$ channels. To obtain an arbitrary desired cycle bound

$$\sum_{j=1}^{m} \tilde{d}_j \leq \sum_{j=1}^{m} (a_{ij} - a_{i,j-1,j})$$

(108)

$∀(i_1, i_2, \ldots, i_m) \in \Pi_M, \forall m \in \{2, 3, \ldots, M\}$, consider the subnetwork consisting of all the receivers and the transmitters $\{i_1, i_2, \ldots, i_m\}$ in Fig. 7, eliminating all other transmitters and their messages. Note that in Fig. 7, for the receivers other than $i_1, i_2, \ldots, i_m$, they are labelled as $i_{m+1}, i_{m+2}, \ldots, i_N$, respectively, and the index ordering of these receivers does not affect the desired outer bound in (109).

Define the message set $W_{θ} ≜ \{W_{ik}, \forall k \in \{m\}$, and the AWGN seen at all the receivers is independent and with zero mean and variance 2.

Then combining with Lemma 1, to obtain the desired cycle bound in (109), we only need to consider the truncated binary-expansion deterministic model of the constructed channel in Fig. 8.

![Fig. 7. An $m \times N$ X channel ($m < N$).](image-url)
For the receivers \( i_j, \forall j \in [m] \), the channel output can be written in the matrix form (110), as shown at the bottom of this page.

While for the virtual receivers, the channel output is given by the following equations

\[
\hat{Y}_{ij}(t) = \begin{bmatrix}
\langle \text{sign}(\bar{X}_{ij}^R(t))h_{R_{ij}}^R & \sum_{b=1}^{m_{ij}} \bar{X}_{ij,b}(t)2^{-b} \rangle \\
\langle \text{sign}(\bar{X}_{ij}^I(t))h_{R_{ij}}^I & \sum_{b=1}^{m_{ij}} \bar{X}_{ij,b}(t)2^{-b} \rangle \\
\end{bmatrix},
\]

\( \forall j \in [m], \forall k \in \{m + 1, m + 2, \ldots, N\} \). (111)

Next, we present a lemma that will be used later.

**Lemma 2**: For any \( j \in [m] \), define \( \hat{S}_j(t) \) and \( \hat{S}_j'(t) \) as (112) and (113), as shown at the bottom of this page, respectively, where the modulo-\( m \) arithmetic is implicitly used on the user indices, e.g., \( i_0 = i_m \). Then \( f : \hat{S}_j(t) \rightarrow \hat{S}_j'(t) \) is bijective.

The proof of Lemma 2 is presented in Appendix B.

Assume that in Fig. 8, \( R_{i_1}^R \) is the strongest virtual receiver connected to transmitter \( i_1 \), i.e., \( |h_{i_1}^R| = \max_{i \in \{m+1, \ldots, N\}} |h_{i_1,i}^R| \). In the truncated binary-expansion deterministic model, the output of receiver \( i_1 \) can be rewritten as (114)-(116), as shown in top of the next page, where \( \hat{C}_{i_1}(t) \) and \( \hat{C}_{i_1}'(t) \) can both take values from \{-1, 0, 1\}. Define

\[
\tilde{X}_{i_1,S}^R(t) = \text{sign}(\bar{X}_{i_1}^R(t)) \sum_{b=1}^{m_{i_1}^R} \bar{X}_{i_1,b}(t)2^{-b}
\]

\( \tilde{X}_{i_1,S}^I(t) = \text{sign}(\bar{X}_{i_1}^I(t)) \sum_{b=1}^{m_{i_1}^I} \bar{X}_{i_1,b}(t)2^{-b} \)

Also define the sum of the terms in (115) and (116) as \( \tilde{Y}_{i_1}(t) \) (see equation (119), as shown in top of the next page).

Then we have

\[
\hat{Y}_{i_1}(t) = \hat{Y}_{i_1}(t) + \hat{Y}_{i_1}(t) + \hat{C}_{i_1}(t)
\]

For the receiver \( i_1 \), starting from Fano’s inequality, we obtain

\[
n(\sum_{j=1}^{m} R_{ij} - \epsilon) \leq I(W_{i_1}^R, \hat{Y}_{i_1}^n, \hat{Y}_{i_1}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C) \]

\[
= H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C) - H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C)
\]

\[
= H(\hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C) + H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C) - H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C)
\]

\[
\leq H(\hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C, \ldots, W_{i_m}^C) + H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C) - H(\hat{Y}_{i_1}^n, \hat{Y}_{i_2}^n, \hat{C}_{i_1}^n, \hat{S}_{i_1}^n | W_{i_1}^C, W_{i_2}^C)
\]

\[
(112)
\]

\[
(113)
\]
\[ \hat{Y}_{i_1}(t) = \left( \frac{\text{sign}(\hat{X}_{i_1}^R(t))R_{i_1}^R \sum_{b=1}^{m_{i_1}^R} \hat{X}_{i_1,b}^R(t) 2^{-b}}{\text{sign}(\hat{X}_{i_1}^L(t))R_{i_1}^R \sum_{b=1}^{m_{i_1}^L} \hat{X}_{i_1,b}^L(t) 2^{-b}} \right) + \left( \frac{\hat{C}_i^R(t)}{\hat{C}_i^L(t)} \right) \]  

(114)

\[ \hat{Y}_{i_1,j}(t) = \left( \frac{\text{sign}(\hat{X}_{i_1}^R(t))R_{i_1}^R \sum_{b=m_{i_1}^R}^{m_{i_1}^R+1} \hat{X}_{i_1,b}^R(t) 2^{-b}}{\text{sign}(\hat{X}_{i_1}^L(t))R_{i_1}^R \sum_{b=m_{i_1}^L}^{m_{i_1}^L+1} \hat{X}_{i_1,b}^L(t) 2^{-b}} \right) + \left( \frac{\hat{C}_i^R(t)}{\hat{C}_i^L(t)} \right) \]  

(115)

\[ \hat{Y}_{i_1}(t) = \left( \frac{\text{sign}(\hat{X}_{i_1}^R(t))R_{i_1}^R \sum_{b=1}^{m_{i_1}^R} \hat{X}_{i_1,b}^R(t) 2^{-b}}{\text{sign}(\hat{X}_{i_1}^L(t))R_{i_1}^R \sum_{b=1}^{m_{i_1}^L} \hat{X}_{i_1,b}^L(t) 2^{-b}} \right) + \left( \frac{\hat{C}_i^R(t)}{\hat{C}_i^L(t)} \right) \]  

(116)

\[ \hat{Y}_{i_1,j}(t) = \left( \frac{\text{sign}(\hat{X}_{i_1}^R(t))R_{i_1}^R \sum_{b=m_{i_1}^R}^{m_{i_1}^R+1} \hat{X}_{i_1,b}^R(t) 2^{-b}}{\text{sign}(\hat{X}_{i_1}^L(t))R_{i_1}^R \sum_{b=m_{i_1}^L}^{m_{i_1}^L+1} \hat{X}_{i_1,b}^L(t) 2^{-b}} \right) + \left( \frac{\hat{C}_i^R(t)}{\hat{C}_i^L(t)} \right) \]  

(117)

where the last inequality follows from the fact that dropping conditioning does not reduce entropy.

Now consider the last term in (124),

\[ H(\hat{Y}_{i_1,n}^n, \hat{Y}_{i_1,n}^n, \hat{C}_{i_1}^n | \hat{W}_{i_2,i_2}^n) \]

(125)

\[ \geq H(\hat{X}_{i_1,1}^R, \hat{X}_{i_1,1}^L, \hat{W}_{i_2,i_2}^n) \]

(126)

\[ = H(\hat{X}_{i_1,1}^R, \hat{X}_{i_1,1}^L, \hat{W}_{i_2,i_2}^n) \]

(127)

Adding (129) and (131), we have

\[ n(\sum_{j=1}^{N} R_{i_1,i} - \sum_{j=m+1}^{N} R_{i_1,j} - 2\epsilon) \]

(132)

Similarly, for \( k \in \{2, 3, \ldots, m-1\} \) we obtain

\[ n(\sum_{j=1}^{N} R_{i_1,i} - \sum_{j=m+1}^{N} R_{i_1,j} - 2\epsilon) \]

(133)

And from the receiver \( i_m \) and the degraded BC comprised of the transmitter \( i_m \) and all its connected virtual receivers, we have

\[ n(\sum_{j=1}^{m} R_{i_m,i} - \sum_{j=m+1}^{N} R_{i_m,i} - 2\epsilon) \]

(134)
Adding all the terms in (132), (133) and (134) together and applying the same argument used in Example 3, we have

\[ n(R_{x,d} - 2\epsilon) \]

\[ \leq \sum_{k=1}^{m} H(\hat{Y}_{ik}^n, \hat{\bar{Y}}_{ik}^n, C_{ik}^n | \hat{S}_k^n) \]

\[ \leq \sum_{k=1}^{m} \sum_{i=1}^{n} \left[ H(\hat{Y}_{ik}^n, \bar{S}_k(t)) + H(\hat{\bar{Y}}_{ik}^n, \bar{S}_k(t)) \right] \]

\[ + H(\hat{C}_i(t)) \]

\[ \leq n \sum_{k=1}^{m} \left[ 2(m_{ik} - m_{i-1,ik}) + \text{constant} \right] \]

\[ \leq n \sum_{k=1}^{m} \left[ (a_{ik} - a_{i-1,ik}) \text{log}_2 P + \text{constant} \right] \]

where (137) holds since \( m_{ik} - 1 \leq \max(m_{ik}^1, m_{ik}^2) \leq m_{ik} \), and also note the modulo-\( m \) arithmetic is implicitly used on user indices, e.g., \( i_0 = i_m \).

Finally according to Lemma 1, we can obtain the desired GDoF cycle bound for the original complex Gaussian \( X \) channel through some simple manipulations,

\[ \sum_{j=1}^{m} \tilde{d}_{ij} \leq \sum_{j=1}^{m} (a_{ij} - a_{i-1,ij}) \]

and hence complete the whole proof.

V. CONCLUSION

In this paper, we extend the optimality of TIN to more general classes of message sets. The main result is that for the TIN-optimal \( K \)-user interference channel, even if the message set expands to include the \( X \) setting where each transmitter has one independent message to each receiver, operating the new channel as the original interference channel and treating interference as noise at each receiver is still optimal to achieve the sum channel capacity to within a constant gap. Furthermore, the optimality of TIN for the general \( M \times N \times X \) channel is also demonstrated.

We conclude with a comment on the necessity of the optimality conditions. In [12] it is conjectured that (8) is also necessary for TIN to be optimal for the entire GDoF region except for a set of channel gain values with measure zero. However, note that no claim is made for the necessity of condition (8) for the optimality of TIN for sum-GDoF. In fact it is easy to see that (8) is not necessary for the sum-GDoF optimality of TIN. For example, consider the 2-user interference channel with \( a_{11} > a_{12} + a_{21} > a_{22} \), which violates (8), and whose optimal sum-GDoF value, \( a_{11} \) (as shown in [7]), is trivially achieved by activating only user 1. It also should be noted that even in the \( 2 \times 2 \times X \) channel, when \( a_{11} > a_{12} + a_{21} > a_{22} \), using the bounding techniques presented in this paper it is not hard to prove that the optimal sum-GDoF value is still \( a_{11} \). Similarly, since our focus is only on sum-GDoF, the optimality conditions are only sufficient, but not necessary.

APPENDIX A

PROOF OF LEMMA 1

We begin with the general complex Gaussian \( M \times N \times X \) channel, and convert it to the corresponding truncated binary-expansion deterministic channel step-by-step. In each step, we show that only a loss of constant bits is introduced. Here we follow the similar steps used in [20]. For the sake of simplicity, we define \( W_i^* \triangleq \{W_{11}, W_{12}, \ldots, W_{1M}\} \), and we suppress the time index \( t \) if no confusion would be caused.

- **Step 1:** Average power constraint to peak power constraint. Recall that in the original complex Gaussian channels, by scaling the output, we set

\[ E[|X_i(t)|^2] \leq 2, \quad Z_i(t) \sim CN(0, 2) \]

Then for each input \( X_i = X_i^R + jX_i^I \), we truncate both the real and imaginary parts to satisfy the peak power constraint of 1. Define the part of input \( X_i^R \) that exceeds the peak power constraint as

\[ \tilde{X}_i^R = |X_i^R| = \text{sign}(X_i^R) \sum_{b=-\infty}^{0} X_{i,b}^R 2^{-b} \]

and the remaining signal as

\[ \tilde{X}_i^R = X_i^R - \tilde{X}_i^R = \text{sign}(X_i^R) \sum_{b=1}^{\infty} X_{i,b}^R 2^{-b} \]

For the imaginary part of the input, we have the similar definitions for \( X_i^I \) with \( I \) replacing \( R \). Then \( \tilde{X}_i^R \) and \( \tilde{X}_i^I \) satisfy the unit peak power constraint. Also define

\[ \tilde{X}_i = \tilde{X}_i^R + j\tilde{X}_i^I \]

\[ \tilde{X}_i = \tilde{X}_i^R + j\tilde{X}_i^I \]

Letting \( \tilde{Y}_i \) be the output of receiver \( i \) due to the truncated input \( \tilde{X}_i \), and \( \hat{Y}_i \) be the difference between \( Y_i \) and \( \tilde{Y}_i \), for each receiver \( i \in [N] \) we have

\[ I(W_i^*; Y_i^n) \leq I(W_i^*; \tilde{Y}_i^n, \tilde{Y}_i^n) \]

\[ = I(W_i^*; \tilde{Y}_i^n) + I(W_i^*; \tilde{Y}_i^n | \tilde{Y}_i^n) \]

\[ \leq I(W_i^*; \tilde{Y}_i^n) + H(\tilde{Y}_i^n) \]

\[ \leq I(W_i^*; \tilde{Y}_i^n) + \sum_{k=1}^{M} H(\tilde{X}_k^R) \]

\[ \leq I(W_i^*; \tilde{Y}_i^n) + n \times \text{constant} \]

The last inequality follows from [20, Lemma 6], which says that for sufficiently large \( n, \) the entropy of \( \tilde{X}_k^R \) (the part of input \( X_k^R \) exceeding the unit peak power constraint) is no larger than a constant number multiplied by \( n \).

- **Step 2:** Truncate signals at noise level and remove noise. Recall \( \log_2[|h_k^R|] = m_k^R \) and \( \log_2[|h_k^I|] = m_k^I \). We have (147) at the top of next page. Next, define \( \tilde{e}_i \) as the difference between \( \tilde{Y}_i \) and \( \hat{Y}_i \) as shown in (148) in next page, where \( \text{frac}(x) \) denotes the fractional part of \( x \). Also note

\[ |h_{ik}^R \sum_{b=m_{ik}^R+1}^{\infty} X_{ik,b}^R 2^{-b}| \leq m_{ik}^R 1 - (m_{ik}^R) = 2 \]

APPENDIX B

PROOF OF LEMMA 1
Finally, we obtain

\[
\hat{Y}_i = \sum_{k=1}^{M} \left[ (\text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}) - (\text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b}) \right] \\
+ j \left[ (\text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}) + (\text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b}) \right]
\]

(147)

\[
\varepsilon_i = \hat{Y}_i - \hat{\hat{Y}}_i
\]

\[
= \sum_{k=1}^{M} \left[ \text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b} - \text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b} \\
+ \text{frac}(\text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}) - \text{frac}(\text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b}) \right]
\]

\[
+ j \left[ \text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b} + \text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b} \\
+ \text{frac}(\text{sign}(X_k^R)h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}) + \text{frac}(\text{sign}(X_k^I)h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b}) \right]
\] + Z_i
\]

(148)

Similarly, we have

\[
|h_{ik}^I \sum_{b=1}^{m_k^I} X_{k,b}^2 2^{-b}| \leq 2
\]

\[
|h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}| \leq 2
\]

\[
|h_{ik}^R \sum_{b=1}^{m_k^R} X_{k,b}^2 2^{-b}| \leq 2
\]

Finally, we obtain

\[
I(W_i^n; \tilde{Y}_i^n)
\leq I(W_i^n; \tilde{Y}_i^n ; e_i^n)
\]

(149)

\[
= I(W_i^n; \tilde{Y}_i^n) + I(e_i^n; \tilde{Y}_i^n | \tilde{Y}_i^n)
\]

(150)

\[
= I(W_i^n; \tilde{Y}_i^n) + h(e_i^n) - h(e_i^n | \tilde{Y}_i^n, W_i^n)
\]

(151)

\[
\leq I(W_i^n; \tilde{Y}_i^n) + h(e_i^n) - h(Z_i^n)
\]

(152)

\[
= I(W_i^n; \tilde{Y}_i^n) + I(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_M; e_i^n)
\]

(153)

\[
\leq I(W_i^n; \tilde{Y}_i^n) + n \times \text{constant}
\]

(154)

where the last inequality is due to the fact that \(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_M \mapsto e_i \) forms a complex Gaussian MAC with a finite SNR independent of \(P \) for each transmitter [20].

\[\square\]

**APPENDIX B**

**PROOF OF LEMMA 2**

For notation brevity, we define

\[
\tilde{X}_R \triangleq \text{sign}(\tilde{X}_i^R(t)) \sum_{b=1}^{m_{ij}^R} \tilde{X}_{i,j,b}^R(t) 2^{-b}
\]

(155)

\[
\max_{m_{ij}^R, m_{ij}^I} \tilde{X}_R \triangleq \text{sign}(\tilde{X}_i^R(t)) \sum_{b=1}^{m_{ij}^R} \tilde{X}_{i,j,b}^R(t) 2^{-b}
\]

(156)

\[
\tilde{S}_R \triangleq |\text{sign}(\tilde{X}_i^R(t))h_{ij}^R | \sum_{b=1}^{m_{ij}^R} \tilde{X}_{i,j,b}^R(t) 2^{-b}
\]

(157)

\[
\hat{\tilde{S}}_R \triangleq \text{sign}(\tilde{X}_i^R(t))h_{ij}^R | \sum_{b=1}^{m_{ij}^R} \tilde{X}_{i,j,b}^R(t) 2^{-b}
\]

(158)

Note (\(\tilde{X}_R, \tilde{X}_I\)) and (\(\hat{\tilde{S}}_R, \hat{\tilde{S}}_I\)) can be seen as the input and output of the deterministic channel, respectively. Obviously, given one input (\(\tilde{X}_R, \tilde{X}_I\)), we can only produce one output (\(\hat{\tilde{S}}_R, \hat{\tilde{S}}_I\)). Next, we prove the other direction by contradiction. We assume there exist two different inputs (\(\tilde{X}_R^a, \tilde{X}_I^a\)) and (\(\tilde{X}_R^b, \tilde{X}_I^b\)) that can generate the same output, i.e., \((\tilde{S}_R^a, \tilde{S}_I^a) = (\tilde{S}_R^b, \tilde{S}_I^b)\). In the following, without loss of generality, we assume

\[
|h_{ij}^R| \geq |h_{ij}^I| \Rightarrow m_{ij}^R \geq m_{ij}^I.
\]

(159)
We first consider the case where \( \text{sign}(h_{ij}^R) = \text{sign}(h_{ij}^L) \). For the term \( \tilde{S}_R \), we have the following subcases:

- \( \tilde{S}_{R,1} = \tilde{S}_{R,2} = \tilde{S}_{R,3} \) and \( \tilde{S}_{R,4} = \tilde{S}_{R,3} \). In this case, for the term \( \tilde{S}_j \), if \( |h_{ij}^R| > |h_{ij}^L| \), since \((\tilde{X}_R^*, \tilde{X}_j^*)\) and \((\tilde{X}_R^+, \tilde{X}_j^+)\) are different, we have \( \tilde{S}_j^* = \tilde{S}_j^+ \) and \( \tilde{S}_j^* \neq \tilde{S}_j^+ \), which contradicts the assumption that \((\tilde{X}_R^*, \tilde{X}_j^*)\) and \((\tilde{X}_R^+, \tilde{X}_j^+)\) generate the same output; if \( |h_{ij}^L| > |h_{ij}^R| \), since \((\tilde{X}_R^+, \tilde{X}_j^+)\) and \((\tilde{X}_R^-, \tilde{X}_j^-)\) generate the same output, we have \((\tilde{X}_R^+, \tilde{X}_j^+) = (\tilde{X}_R^-, \tilde{X}_j^-)\), which contradicts the assumption that \((\tilde{X}_R^*, \tilde{X}_j^*)\) and \((\tilde{X}_R^+, \tilde{X}_j^+)\) generate the same output.

- \( \tilde{S}_{R,1} > \tilde{S}_{R,2} > \tilde{S}_{R,3} \). In this case, for the term \( \tilde{S}_j \), we have \( \tilde{S}_{j,1} \geq \tilde{S}_{j,1}^* \) and \( \tilde{S}_{j,2} > \tilde{S}_{j,2}^* \), which contradicts the assumption that \((\tilde{X}_R^*, \tilde{X}_j^*)\) and \((\tilde{X}_R^+, \tilde{X}_j^+)\) generate the same output.

- \( \tilde{S}_{R,1} < \tilde{S}_{R,2} < \tilde{S}_{R,3} \). In this case, for the term \( \tilde{S}_j \), we have \( \tilde{S}_{j,1} \leq \tilde{S}_{j,1}^* \) and \( \tilde{S}_{j,2} < \tilde{S}_{j,2}^* \), which contradicts the assumption that \((\tilde{X}_R^+, \tilde{X}_j^+)\) and \((\tilde{X}_R^-, \tilde{X}_j^-)\) generate the same output.

For the other case where \( \text{sign}(h_{ij}^R) = -\text{sign}(h_{ij}^L) \), we can follow the above argument to get the same conclusion. \( \square \)

**Acknowledgment**

The authors would like to thank the Associate Editor, and the reviewers for their valuable and timely comments.

**References**


