On the Generalized Degrees of Freedom of the Noncoherent Interference Channel

Joyson Sebastian and Suhas Diggavi, *Fellow, IEEE*

***Abstract*— We study the generalized degrees of freedom (gDoF) of the block-fading *noncoherent* 2-user interference channel (IC) with a coherence time of T symbol durations and symmetric fading statistics. We demonstrate that a standard training-based scheme for the noncoherent IC is suboptimal in several regimes. We study and analyze several alternate schemes: the first is a new noncoherent scheme using rate-splitting. We also consider a scheme that treats interference-as-noise (TIN) and a time division multiplexing (TDM) scheme. We show that a standard training-based scheme for the noncoherent IC is outperformed by one of these schemes in several regimes: our results demonstrate that in the very weak interference regime, the TIN scheme is the best; in the strong interference regime, the TDM scheme and the noncoherent rate-splitting scheme give better performance; in other cases either of the TIN, TDM or noncoherent rate-splitting scheme could be preferred. We also study the noncoherent IC with feedback and propose another noncoherent rate-splitting scheme. Again for the feedback case, our results demonstrate that a standard training-based scheme can be outperformed by other schemes.**

***Index Terms*— Noncoherent communication, degrees of freedom (DoF), time-varying channels, interference channels, channels with feedback.**

I. Introduction

NONCOHERENT wireless channels where neither the transmitter nor the receiver knows the channel [1]–[5] have been studied for point-to-point communication systems. To the best of our knowledge, the *noncoherent* interference channel (IC) has not been studied from an information theoretic viewpoint. In this paper, we consider the noncoherent 2-user IC with symmetric statistics and study the generalized degrees of freedom (gDoF) region as a first step towards understanding its capacity region.

A. Related Work

To the best of our knowledge, the capacity of the noncoherent interference channel has not received much attention in the literature. Hence, we give an overview of the existing works on noncoherent wireless networks and the related work on the interference channels. The noncoherent wireless model for the multiple-input multiple-output (MIMO) channel was studied by Marzetta and Hochwald [1]. In their model, neither the receiver nor the transmitter knows the fading coefficients and the fading gains remain constant within a block of length T symbol periods. Across the blocks, the fading gains are independent and identically distributed (i.i.d.) according to a Rayleigh distribution. The capacity behavior at high average signal-to-noise ratio[1] (SNR) for the noncoherent MIMO channel was studied by Zheng and Tse in [3]. The main conclusion of that work was that a standard training-based scheme was DoF optimal for the noncoherent MIMO channels, a message distinct from our conclusions in this paper for the noncoherent IC. Some works have specifically studied the case with $T = 1$ [2], [6], [7]. In [2], it was demonstrated that for $T = 1$, the capacity is achieved by a distribution with a finite number of mass points, but the number of mass points grows with the SNR. The capacity for the case with $T = 1$ was shown to behave double-logarithmically in [7].

There have been other works that studied noncoherent relay channels. The noncoherent single relay network was studied in [4], where the authors considered identical link strengths and unit coherence time. They showed that under certain conditions on the fading statistics, the relay does not increase the capacity at high SNR. In [8], similar observations were made for the noncoherent MIMO full-duplex single relay channel with block-fading. The authors showed that Grassmanian signaling can achieve the DoF without using the relay. Also for certain regimes, decode-and-forward with Grassmanian signaling was shown to approximately achieve the capacity at high SNR.

The above works considered a DoF framework for the noncoherent model in the sense that for high SNR, the link strengths are not significantly different, *i.e.,* the links scale with the same SNR-exponent. In contrast, a gDoF framework considers the situation when the link strengths could have large difference. The gDoF framework for the noncoherent MIMO channel was considered in [9], [10] and it was shown that several insights from the DoF framework may not carry on to the gDoF framework. It was shown that a standard training-based scheme is not gDoF optimal and that all antennas may have to be used for achieving the gDoF, even when the coherence time is low, in contrast to the results for the MIMO channel with i.i.d. links. In [5], the gDoF of the 2-relay

Manuscript received 21 December 2020; revised 7 May 2021 and 2 November 2021; accepted 2 February 2022. Date of publication 2 March 2022; date of current version 12 September 2022. This work was supported in part by NSF Grant 1955632 and Grant 2007714 and in part by the Gift from Guru Krupa Foundation. The associate editor coordinating the review of this article and approving it for publication was M. C. Gursoy. *(Corresponding author: Joyson Sebastian.)*

Joyson Sebastian was with the Department of Electrical and Computer Engineering (ECE), University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA. He is now with Samsung, San Diego, CA 92126 USA (e-mail: joysonsebastian@ucla.edu).

Suhas Diggavi is with the Department of Electrical and Computer Engineering (ECE), University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA (e-mail: suhasdiggavi@ucla.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3153893.

Digital Object Identifier 10.1109/TWC.2022.3153893

[1]We use the abbreviation SNR for the *average* signal-to-noise ratio in the context of fading channels and not for the (instantaneous) signal-to-noise ratio. Similarly, we use the abbreviation INR for the *average* interference-to-noise ratio.

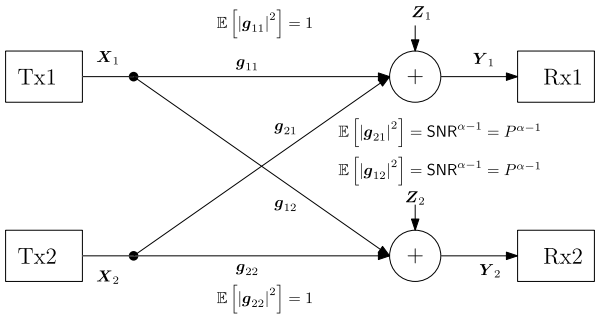

Fig. 1. The channel model without feedback.

diamond network was studied. The standard training-based schemes were proved to be sub-optimal and a new scheme was proposed, which partially trains the network and performs a scaling and quantize-map-forward operation [11]–[13] at the relays.

In this work, we study the noncoherent 2-user IC with symmetric statistics. This, we believe, is the first information theoretic analysis of noncoherent channels in multiple unicast networks with interference. The capacity of the (coherent) 2-user Gaussian IC is well studied [14]–[17] when the channels are perfectly known at the receivers and transmitters. The capacity region of the 2-user IC without feedback was characterized in [16], to within 1 bit per user. In [17], a similar result was derived for the 2-user Gaussian IC with feedback, obtaining the capacity region within 2 bits per user. In [18], the approximate capacity region (within a constant additive gap) for 2-user fast fading interference channels (FF-IC), with no instantaneous CSIT but with perfect channel knowledge at the receiver, was derived. There, the authors used a rate-splitting scheme based on the average interference-to-noise ratio, extending the existing rate-splitting schemes for the IC [16], [17]. The approximate capacity region was derived for the FF-IC without feedback and also for the case with feedback; the feedback improves the capacity region for the FF-IC, similar to the case for the static IC [17]. In this work, we extend the results from [18] for the FF-IC (where the receivers know the channel, but not the transmitters) to the case when both transmitters and receivers do not know the channel, *i.e.,* the noncoherent IC.

B. System Model, Performance Metrics, and Contributions

Our system model is illustrated in Figure 1. We have two transmitters, each with its own intended receiver. The transmitted signals are multiplied by random fading noise. Each receiver receives a sum of signals from both transmitters and with additive white Gaussian noise (AWGN). In our model, the additive white Gaussian noise and the direct channel links are normalized to be of unit power and hence the power P at transmitters is set to be equal to the SNR, *i.e.* $P=\mathsf{SNR}$. With the AWGN at unit power level, the average interference-to-noise ratio (INR) is the average received power through the interfering links $P\times\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right], P\times\mathbb{E}\left[|\boldsymbol{g}_{21}|^2\right]$, *i.e.* $P\times\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right]=P\times\mathbb{E}\left[|\boldsymbol{g}_{21}|^2\right]=\mathsf{INR}$. The interfering links are set to scale as $\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right]=\mathbb{E}\left[|\boldsymbol{g}_{21}|^2\right]=P^{\alpha-1}$, where the parameter α (interference level) is used to capture the relative strength of the interference at the receiver:

$$\alpha=\frac{\log(\mathsf{INR})}{\log(\mathsf{SNR})}=\frac{\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{21}|^2\right]\right)}{\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{11}|^2\right]\right)}=\frac{\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right]\right)}{\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{22}|^2\right]\right)}=\frac{\log\left(P\times P^{\alpha-1}\right)}{\log\left(P\times 1\right)}.$$

Typically the interference level is such that $\mathsf{INR}<\mathsf{SNR}$. When $\alpha<1/2$, we have the very weak interference regime. When $1/2\le\alpha\le 1$, we have the weak interference regime. We also consider the strong interference regime[2] ($\alpha>1$) where the interfering links are stronger than the direct links. For example for a case with transmit power 23 dB, direct links with average strength 0.1 and crosslinks with average strength 0.2, the interference level is $\log\left(10^{23/10}\times.2^2\right)/\log\left(10^{23/10}\times.1^2\right)=3$. For a case with transmit power 25 dB, direct links with average strength 0.1 and crosslinks with average strength 0.32, the interference level is again $\log\left(10^{25/10}\times.32^2\right)/\log\left(10^{25/10}\times.1^2\right)=3$. Therefore the α same translates to different channel strengths at different SNR.

For the channel, we consider a block fading model where the channels remain constant for a coherence time of T symbol durations. Thus we model our system with vectors of size T. We have

$$\boldsymbol{Y}_1=\boldsymbol{g}_{11}\boldsymbol{X}_1+\boldsymbol{g}_{21}\boldsymbol{X}_2+\boldsymbol{Z}_1, \tag{1}$$

$$\boldsymbol{Y}_2=\boldsymbol{g}_{12}\boldsymbol{X}_1+\boldsymbol{g}_{22}\boldsymbol{X}_2+\boldsymbol{Z}_2, \tag{2}$$

where the $\boldsymbol{X}_i$, $\boldsymbol{Y}_i$, $\boldsymbol{Z}_i$ with $i\in\{1,2\}$ are transmitted symbols, received symbols and noise at receivers respectively. The variables $\boldsymbol{X}_i$, $\boldsymbol{Y}_i$, $\boldsymbol{Z}_i$ with $i\in\{1,2\}$ are $1\times T$ vectors. We have the average power constraint on the transmitted signals

$$\frac{1}{T}\mathbb{E}\left[|\boldsymbol{X}_i|^2\right]=P=\mathsf{SNR} \tag{3}$$

for $i\in\{1,2\}$. The noise $\boldsymbol{Z}_1,\boldsymbol{Z}_2$ are independent of each other and their realizations are i.i.d. across time. The entries of the vector $\boldsymbol{Z}_i$ are i.i.d. according to $\mathcal{CN}(0,1)$. The fading channels are indicated by scalar random variables $\boldsymbol{g}_{ij}$ with $i,j\in\{1,2\}$. The realizations of $\boldsymbol{g}_{ij}$ for any fixed (i,j), $i,j\in\{1,2\}$ are i.i.d. across time, and the realizations for different (i,j) are independent. We consider the case with symmetric fading statistics $\boldsymbol{g}_{11}\sim\boldsymbol{g}_{22}\sim\mathcal{CN}(0,1)$, $\boldsymbol{g}_{12}\sim\boldsymbol{g}_{21}\sim\mathcal{CN}\left(0,P^{\alpha-1}\right)$. Neither the receivers nor the transmitters have knowledge of any of the realizations of $\boldsymbol{g}_{ij}$, but the channel statistics are known to all the receivers and transmitters.

[2]A different way of looking at the scaling in the strong interference regime ($\alpha>1$) is to use a different normalization $\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right]=\mathbb{E}\left[|\boldsymbol{g}_{21}|^2\right]=1$, $P=\mathsf{INR}$, $\mathbb{E}\left[|\boldsymbol{g}_{11}|^2\right]=\mathbb{E}\left[|\boldsymbol{g}_{22}|^2\right]=P^{1/\alpha-1}$. Thus we still have $\alpha=\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{12}|^2\right]\right)/\log\left(P\times\mathbb{E}\left[|\boldsymbol{g}_{11}|^2\right]\right)=\log(\mathsf{INR})/\log(\mathsf{SNR})$ capturing the relative strengths, the transmit power P is scaled, and the strengths of direct and interfering links remain bounded.

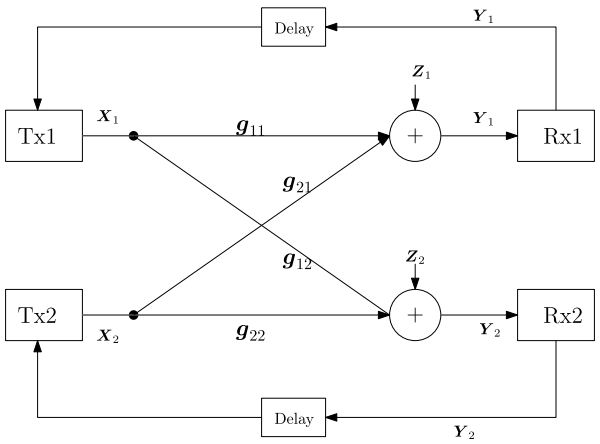

Fig. 2. The channel model with feedback.

We also consider a feedback model (Figure 2), where each receiver reliably feeds back the received symbols[3] to the corresponding transmitter. We consider the feedback of symbols in blocks of T: the symbols are fed back after all the symbols in a coherence period of T are received. However, the results that we derive are valid even if the feedback is performed after receiving each symbol.

The main metric used in this paper to evaluate the performance of achievability schemes is the gDoF. It is desirable to have an exact capacity characterization, but this is unknown even for the coherent IC except for some regimes [16], [20]. For the noncoherent case, the exact capacity characterization is open even for point-to-point channels. Hence, we give a characterization for the noncoherent IC in terms of the asymptotic approximation of the capacity region defined by the gDoF region. For this, we consider a series of channels with a fixed interference level α and letting $\mathsf{SNR},\ \mathsf{INR} \to \infty$. Let $\mathcal{C}(\mathsf{SNR},\mathsf{INR})$ denote the capacity region of the channel and let $\tilde{\mathcal{D}}_C$ be a scaled version of $\mathcal{C}(\mathsf{SNR},\mathsf{INR})$ given by $\tilde{\mathcal{D}}_C(\mathsf{SNR},\mathsf{INR}) = \{(R_1/\log(\mathsf{SNR}), R_2/\log(\mathsf{SNR})) : (R_1,R_2) \in \mathcal{C}(\mathsf{SNR},\mathsf{INR})\}$. Following [16], we define the generalized degrees of freedom region as the asymptote of the scaled capacity region:

$$\mathcal{D}_C(\alpha) = \lim_{\substack{\mathsf{SNR},\ \mathsf{INR}\to\infty \\ \alpha \text{ fixed}}} \tilde{\mathcal{D}}_C(\mathsf{SNR},\mathsf{INR}). \tag{4}$$

In other words, $\mathcal{D}_C(\alpha)$ contains elements (d_1,d_2) iff (d_1,d_2) lies within $\tilde{\mathcal{D}}_C(\mathsf{SNR},\mathsf{INR})$ in the asymptotic case of $\mathsf{SNR},\ \mathsf{INR} \to \infty$ with fixed α. This can be formally stated as:

$$\mathcal{D}_C(\alpha) = \Big\{ (d_1,d_2) : \lim_{\substack{\mathsf{SNR},\ \mathsf{INR}\to\infty \\ \alpha \text{ fixed}}} \times \Big(\min_{(y_1,y_2)\in\tilde{\mathcal{D}}_C(\mathsf{SNR},\mathsf{INR})} |(d_1,d_2)-(y_1,y_2)| \Big) = 0 \Big\} \tag{5}$$

[3]The IC with rate limited feedback is considered in [19] where outputs are quantized and fed back. Our schemes can also be extended for such cases.

If we have any rate region $\mathcal{R}(\mathsf{SNR},\mathsf{INR})$, we can similarly define a prelog region $\mathcal{D}_R(\alpha)$ in the following manner:

$$\mathcal{D}_C(\alpha) = \Big\{ (R_1/\log(\mathsf{SNR}), R_2/\log(\mathsf{SNR})) : (R_1,R_2) \in \mathcal{R}(\mathsf{SNR},\mathsf{INR}) \Big\} \tag{6}$$

$$\mathcal{D}_C(\alpha) = \Big\{ (d_1,d_2) : \lim_{\substack{\mathsf{SNR},\ \mathsf{INR}\to\infty \\ \alpha \text{ fixed}}} \times \Big(\min_{(y_1,y_2)\in\tilde{\mathcal{D}}_R(\mathsf{SNR},\mathsf{INR})} |(d_1,d_2)-(y_1,y_2)| \Big) = 0 \Big\} \tag{7}$$

An achievable prelog region is the prelog region derived from an achievable rate region. Just as the capacity region is the maximal achievable rate region, so the gDoF region is the maximal achievable prelog region.[4] Our scaling process with the parameter α creates a sequence of channels indexed by SNR. If a sequence of channels is not created, but merely the transmit power is scaled, we are able to capture the asymptotic behaviour of the capacity region $\mathcal{C}(\mathsf{SNR},\mathsf{INR})$ along the line $\log(\mathsf{SNR}) = \log(\mathsf{INR}) \to \infty$ in the 2-dimensional space of the parameters $\log(\mathsf{SNR}),\ \log(\mathsf{INR})$. With the sequence of channels as created in our setting, we capture the asymptotic behaviour of the capacity region along a general line $\alpha\log(\mathsf{SNR}) = \log(\mathsf{INR}) \to \infty$. This gives us a better understanding of the nature of the capacity region, when the exact capacity region itself is unknown.

This method of characterization was first used in [16] to characterize the asymptotic behavior of the capacity region of a 2-user symmetric IC for high SNR. In [16], the received signal strengths through the four links of the IC were set to scale as $\mathsf{SNR}, \mathsf{SNR}^{\alpha}, \mathsf{SNR}^{\alpha}, \mathsf{SNR}$. The method of scaling the received signal strengths on different links with different SNR-exponents to obtain the gDoF region is also used in other works like [5], [10], [21], [22]. A similar way of creating a sequence of channels depending on SNR is also found in diversity multiplexing trade-off formulation [23]. The high-SNR capacity of noncoherent networks scale as $\log(\log(\mathsf{SNR}))$ for $T=1$, following the result for the MIMO channel [6], [7], [10]; as a consequence the gDoF region is null for $T=1$, and therefore we assume $T \geq 2$ in our paper.

A standard training-based scheme estimates the channel at the receiver using known training symbols sent from the transmitter and uses the estimate to operate a coherent decoder. Such a scheme is known to be DoF optimal for the noncoherent single-user MIMO channel [3]. A natural question to ask is whether operating the noncoherent IC with such a standard training-based scheme achieves the gDoF region. The main observation in this paper is that we can improve the prelog region of the standard training-based coherent schemes in several regimes for the noncoherent IC.

We provide the coding schemes and analysis for the following schemes.

1) We develop a noncoherent version of the simplified Han-Kobayashi scheme from [16] for the 2-user IC, where

[4]For point-to-point channels, the prelog of an achievable rate R from a scheme can be defined as $\lim_{\mathsf{SNR}\to\infty} R/\log(\mathsf{SNR})$ and the DoF is the maximal achievable prelog.

the transmitters use superposition coding, rate-splitting their messages into common and private parts based on the INR. Each receiver noncoherently decodes its own private message and the common messages from both users.

2) Similar to the previous scheme, another noncoherent scheme is developed for the 2-user IC with feedback extending the coherent scheme from [17]. This scheme involves B blocks. In the first block, each transmitter splits its own message into common and private parts and then sends a codeword superimposing the common and private messages. In subsequent blocks, the common message from the other user is decoded at the transmitter using the feedback. Each transmitter generates new common and private messages, conditioned on the previous common messages from both users. After a total of B blocks, each receiver performs backward decoding. Each decoding step in this scheme is performed noncoherently.
3) A training-based scheme is analyzed for the noncoherent IC without feedback. The first two symbols in every coherence period of T symbols is used for estimating the channels.[5] The rest of the symbols are used for transmitting data. The data transmission is performed according to a rate-splitting scheme [16] for the coherent IC: the transmitters use superposition coding, rate-splitting their messages into common and private parts based on the INR. Each receiver uses the channel estimates and decodes its own private message and the common messages from both users.
4) A training-based scheme is analyzed for the noncoherent IC with feedback. The first two symbols in every coherence period of T symbols is used for estimating the channels. The rest of the symbols are used for transmitting data. The data transmission is performed according to a rate-splitting scheme [17] for the coherent IC with feedback. This is similar to the scheme 2) that we described for the noncoherent case, except that the decoding is performed coherently using the estimated channel values.
5) We consider a scheme which treats interference-as-noise (TIN) where each receiver treats the symbols from the other user as noise. The first symbol in every coherence period is used for estimating the channels. Each user estimates its own channel while treating the other user as interference.
6) We also consider a time division multiplexing (TDM) between single-user transmissions with equal time-sharing between the users. Alternate blocks of length T are used by alternate users. For each user, the first symbol in the block of length T is used for estimating its channel.

The TIN and TDM schemes are implemented using one training symbol in each coherence period, as there is only one channel coefficient to be estimated for each user. The TIN and TDM schemes can also be implemented in a noncoherent manner without training symbols, but it can be verified that the prelog performance remains the same. We evaluate the achievable prelog region with the above schemes and compare the performance. Our main results on the prelog of the noncoherent IC are illustrated in Figure 3 and Figure 4.

[5]As we are considering high-SNR behavior, one training symbol is sufficient for each link.

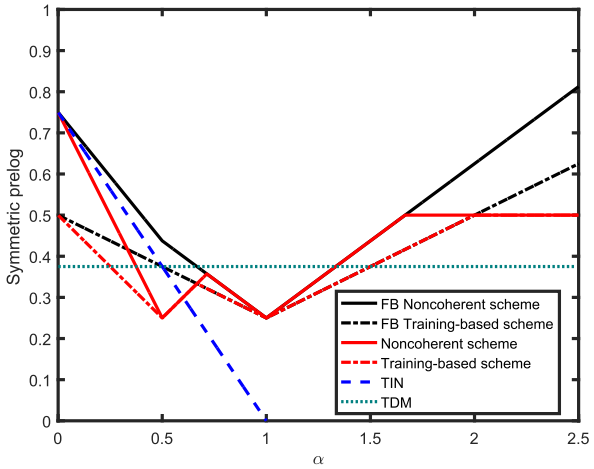

Fig. 3. Symmetric achievable prelog of the noncoherent IC for coherence time $T = 4$.

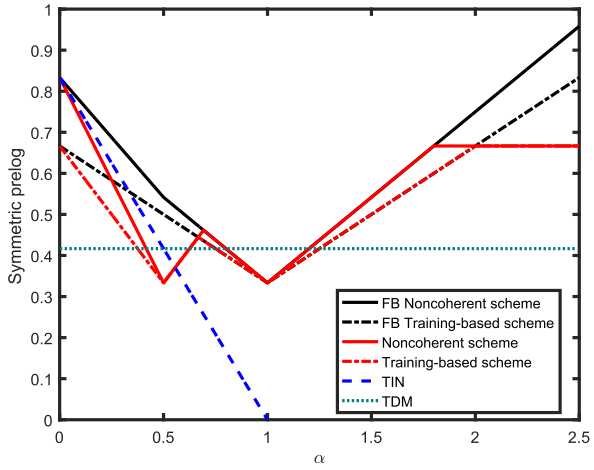

Fig. 4. Symmetric achievable prelog of the noncoherent IC for coherence time $T = 6$.

When the INR is much lower than the SNR in the absence of feedback, the TIN scheme is better than other schemes that decode part of the interfering message. In contrast, for the case when the channel is perfectly known, the TIN scheme has the same performance as a rate-splitting scheme without feedback when the INR is much lower than the SNR. However, for the noncoherent case, rate-splitting schemes without feedback have lower prelog. We believe that this is due to the added uncertainty in the interfering link along with the uncertainty of the interfering message to be decoded. Due to this added uncertainty, it also is better to avoid interference using the TDM scheme when the interference level α is close to 1.

In general, the noncoherent schemes perform better than the standard training-based schemes. The schemes with feedback have larger prelog than the corresponding schemes without

TABLE I
ACHIEVABLE PRELOG REGION FOR DIFFERENT REGIMES OF α

$\alpha < 1/2$	$1/2 \leq \alpha \leq 1$	$\alpha > 1$
$d_1 \leq (1-1/T) - \alpha/T$ $d_2 \leq (1-1/T) - \alpha/T$ $d_1 + d_2 \leq 2(1-1/T) - 2\alpha$	$d_1 + d_2 \leq (2-3/T) - \alpha(1-1/T)$ $d_1 + d_2 \leq 2(1-2/T)\alpha$ $2d_1 + d_2 \leq (2-3/T) - \alpha/T$ $d_1 + 2d_2 \leq (2-3/T) - \alpha/T$	$d_1 \leq (1-2/T)$ $d_2 \leq (1-2/T)$ $d_1 + d_2 \leq (1-1/T)\alpha - 1/T$

feedback. With feedback, the performance of noncoherent rate-splitting schemes is in general better than the TIN scheme. However, the TDM scheme is still the best around $\alpha = 1$.

We also provide some numerical results on achievable rates to show that our gDoF results can provide improvements in the rates compared to the standard training-based schemes at finite SNRs, the rate-SNR points are given in Table III on page 7.

The rest of the paper is organized as follows. In Section I-C, we explain the notations used. In Section II, we discuss our results on the noncoherent IC without feedback and in Section III, we discuss the noncoherent IC with feedback. In Section IV, we give the conclusions and remarks. Some of the proofs for the analysis are deferred to the appendices.

C. Notational Conventions

We use the notation $\mathcal{CN}\left(\mu, \sigma^2\right)$ for circularly symmetric complex Gaussian distribution with mean μ and variance σ^2. The logarithm to base 2 is denoted by $\log()$. We use the symbol $\sim$ with overloaded meanings: one to indicate that a random variable has a given distribution and second to indicate that two random variables have the same distribution. We use the notation $\doteq$ for order equality, *i.e.*, we say $f_1(\mathsf{SNR}) \doteq f_2(\mathsf{SNR})$ if

$$\lim_{\mathsf{SNR}\to\infty} \frac{f_1(\mathsf{SNR})}{\log(\mathsf{SNR})} = \lim_{\mathsf{SNR}\to\infty} \frac{f_2(\mathsf{SNR})}{\log(\mathsf{SNR})}. \tag{8}$$

The use of symbols $\dot{\leq}, \dot{\geq}$ are defined analogously. When we have an SNR-dependent term t_1 in evaluating the rate of a scheme, we have the prelog of the term t_1 as the limit $\lim_{\mathsf{SNR}\to\infty} t_1/\log(\mathsf{SNR})$. We use a bold script for random variables and the normal script for deterministic variables. We use small letters for scalars, capital letters for vectors and capital letter with underline for matrices. The following capital letters being a common notation are used for scalars: P for power, B for number of codeblocks, T for the coherence time, R for rate and C for capacity. The special script of the form $\mathcal{A}, \mathcal{E}$ is used to indicate sets. As an exception to this rule, the notation $\mathcal{CN}$ is reserved for complex Gaussian distribution. The notation $\boldsymbol{x}_{i,j}$ indicates the j^{th} element of $\boldsymbol{X}_i$. Similar definitions follow for $\boldsymbol{y}_{i,j}$ and $\boldsymbol{z}_{i,j}$. The random variables $\boldsymbol{g}_{ij}$ with $i, j \in \{1, 2\}$ are scalar random variables to capture the block fading.

II. NONCOHERENT IC WITHOUT FEEDBACK

In this section, we provide our results for the noncoherent IC without feedback. We compare the achievable prelog using a standard training-based scheme to our noncoherent rate-splitting scheme and we also compare it with the TIN and TDM schemes.

Theorem 1: Using a noncoherent rate-splitting scheme, the prelog region given in Table I is achievable.

Proof: The proof follows by analyzing a Han-Kobayashi scheme [14], [15] with rate-splitting based on the average interference-to-noise ratio [18]. The message for User 1 is split into two parts, a common message $w_{\mathrm{c}1}$ at rate $R_{\mathrm{c}1}$ and a private message $w_{\mathrm{p}1}$ at rate $R_{\mathrm{p}1}$. The common message $w_{\mathrm{c}1}$ is mapped into Gaussian vector symbols represented by $\boldsymbol{U}_1$ and private message $w_{\mathrm{p}1}$ is mapped into Gaussian vector symbols represented by $\boldsymbol{X}_{\mathrm{p}1}$ where $\boldsymbol{U}_1, \boldsymbol{X}_{\mathrm{p}1}$ are independent. The vectors are of size T. The transmitted symbols at Transmitter 1 are of the form $\boldsymbol{X}_1 = \boldsymbol{U}_1 + \boldsymbol{X}_{\mathrm{p}1}$. The power allocation to the symbols are determined based on the average interference-to-noise ratio. The power of each element of $\boldsymbol{X}_{\mathrm{p}1}$ is $P^{1-\alpha}$ and the power of each element of $\boldsymbol{U}_1$ is $P - P^{1-\alpha}$.

Similarly at Transmitter 2, we have a common message $w_{\mathrm{c}2}$ at rate $R_{\mathrm{c}2}$ and a private message $w_{\mathrm{p}2}$ at rate $R_{\mathrm{p}2}$. The common message $w_{\mathrm{c}2}$ is mapped into Gaussian vector symbols represented by $\boldsymbol{U}_2$ and private message is mapped into Gaussian vector symbols represented by $\boldsymbol{X}_{\mathrm{p}2}$, where $\boldsymbol{U}_2, \boldsymbol{X}_{\mathrm{p}2}$ are independent. The transmitted symbols at Transmitter 2 are of the form $\boldsymbol{X}_2 = \boldsymbol{U}_2 + \boldsymbol{X}_{\mathrm{p}2}$. The power of each element of $\boldsymbol{X}_{\mathrm{p}2}$ is $P^{1-\alpha}$ and the power of each element of $\boldsymbol{U}_2$ is $P - P^{1-\alpha}$.

Each receiver, in a noncoherent manner jointly decodes its own private message and the common messages from both users, *i.e.*, Receiver 1 decodes $w_{\mathrm{c}1}, w_{\mathrm{c}2}, w_{\mathrm{p}1}$ and Receiver 2 decodes $w_{\mathrm{c}1}, w_{\mathrm{c}2}, w_{\mathrm{p}2}$. The details of the coding scheme and its analysis are in Section II-B. ■

We now compare our achievable prelog with that of a standard training-based scheme.

Theorem 2: A standard training-based scheme for the noncoherent IC can achieve the prelog region described in Table II.

Proof: With two users, in every coherence period of T symbols, we need at least two symbols for training. For training, the first transmitter can send a known symbol while the second transmitter remains turned off. With this, both receivers can estimate the channels from the first transmitter. Next the second transmitter can send a known symbol while the first transmitter remains turned off. With this, both receivers can estimate the channels from the second transmitter. The rest of the symbols in every coherence period of T symbols can be used to transmit data using a Han-Kobayashi scheme scheme similar to that described in Theorem 1. The detailed analysis for obtaining the prelog is given in [24, Appendix C]. ■

TABLE II
ACHIEVABLE PRELOG REGION FOR DIFFERENT REGIMES OF α

$\alpha < 1/2$	$1/2 \leq \alpha \leq 1$	$\alpha > 1$
$d_1 \leq (1-2/T)$ $d_2 \leq (1-2/T)$ $d_1+d_2 \leq 2(1-2/T)(1-\alpha)$	$d_1+d_2 \leq (1-2/T)(2-\alpha)$ $d_1+d_2 \leq 2(1-2/T)\alpha$ $2d_1+d_2 \leq 2(1-2/T)$ $d_1+2d_2 \leq 2(1-2/T)$	$d_1 \leq (1-2/T)$ $d_2 \leq (1-2/T)$ $d_1+d_2 \leq (1-2/T)\alpha$

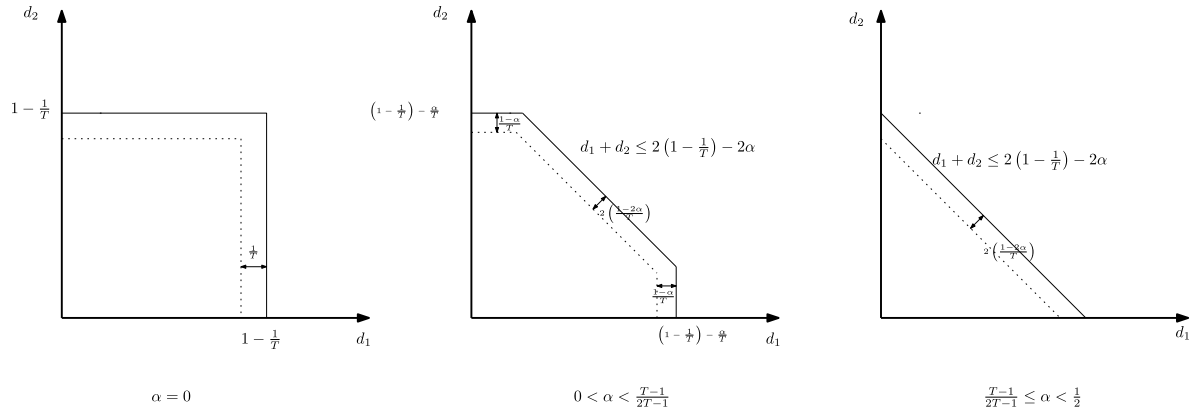

Fig. 5. Prelog region for $\alpha < 1/2$, $T \geq 2$. The solid line gives the prelog region achievable for a noncoherent scheme and the dotted line gives the prelog for the scheme that uses 2 symbols for training.

Remark 1: The capacity region of the coherent FF-IC is known within a constant gap from [18] and hence its gDoF region is known. The prelog region from the above theorem is the same as the gDoF region for the coherent FF-IC with a multiplication factor of $(1-2/T)$. *Hence the prelog region obtained in Theorem 2 is the best among any scheme that uses two symbols for training in every coherence period of* T *symbols.*

We also consider the strategy of treating-interference-as-noise (TIN) with Gaussian codebooks. Transmitter i sends a message w_i at rate R_i using vector Gaussian symbols $\boldsymbol{X}_i$ of length T, $i \in \{1,2\}$. Receiver i decodes w_i, treating the symbols from the other transmitter as noise for $i \in \{1,2\}$. Using standard analysis, it can be shown that the prelog region

$$d_1 \leq (1-1/T)(1-\alpha) \tag{9a}$$
$$d_2 \leq (1-1/T)(1-\alpha) \tag{9b}$$

is achievable by the TIN scheme.

Another strategy is time division multiplexing (TDM). Again Transmitter i can send a message w_i at rate R_i using vector Gaussian symbols $\boldsymbol{X}_i$ of length T, $i \in \{1,2\}$. For the TDM case, each transmitter transmits in every alternate time periods of length T. While one transmitter is ON, the other is OFF. Each receiver obtains symbols only from the intended transmitter and can perform typicality decoding. Using standard analysis it can be shown that the prelog region

$$d_1 \leq (1/2)(1-1/T), \tag{10a}$$
$$d_2 \leq (1/2)(1-1/T), \tag{10b}$$

is achievable.

A. Discussion

In Figure 5 and Figure 6, the prelog region achievable using our noncoherent scheme is compared with the prelog region achievable using the aforementioned training-based scheme. It can be observed that our noncoherent scheme outperforms the standard training-based scheme.

In Figure 7 and Figure 8, we give the achievable symmetric prelog with coherence time $T = 3$ and $T = 5$ respectively for the strategies that we discussed. In the overview of our results in Section I, we had noticed that TIN outperforms rate-splitting schemes. In fact, it can be calculated from our prelog regions that the TIN scheme also outperforms the TDM scheme for very weak interference level ($\alpha < 1/2$).

For a broad region of α, the TDM scheme outperforms the noncoherent rate-splitting scheme. This can be clearly seen by looking at the points with $\alpha = .5$ and $\alpha = 1$. For these values of α, the noncoherent rate-splitting scheme gives a prelog of $(1/2)(1-2/T)$ and the TDM scheme gives a prelog of $(1/2)(1-1/T)$. Hence, for $\alpha = .5$ and $\alpha = 1$, the noncoherent scheme effectively behaves as a TDM scheme that uses two training symbols per coherence period, where actually the TDM scheme can be implemented with only one training symbol per coherence period.

Although our main results are on the prelog of the system, our analysis can show that some schemes are more efficient than others in terms of achievable rates for certain values of power. For example with transmit SNR 6 dB, coherence time $T = 5$, and all the links with average strength 1, using the TDM scheme can improve the rate by 6% compared to the standard training-based schemes used with rate-splitting. More rate points are illustrated in Table III. The rates for training-based scheme is obtained by numerically evaluating

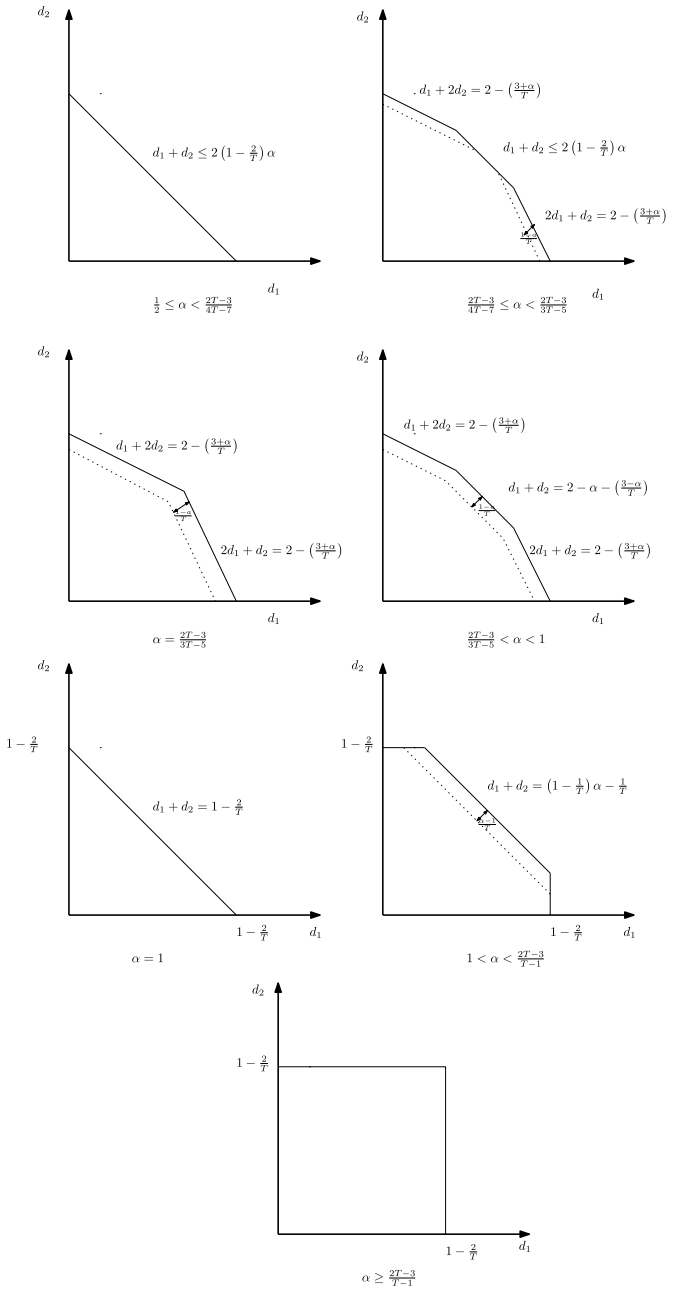

Fig. 6. Prelog region for $1/2 \le \alpha$, $T \ge 3$. For $T = 2$ no prelog region is achievable using known schemes. The solid line gives the prelog region achievable for a noncoherent scheme and the dotted line gives the prelog region for a scheme that uses 2 symbols for training.

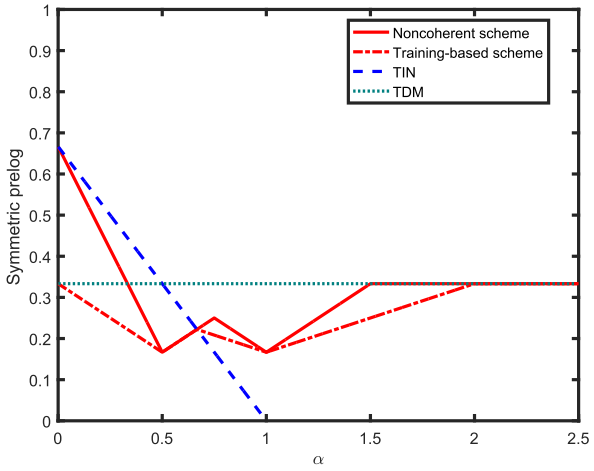

Fig. 7. Symmetric achievable prelog for coherence time $T = 3$.

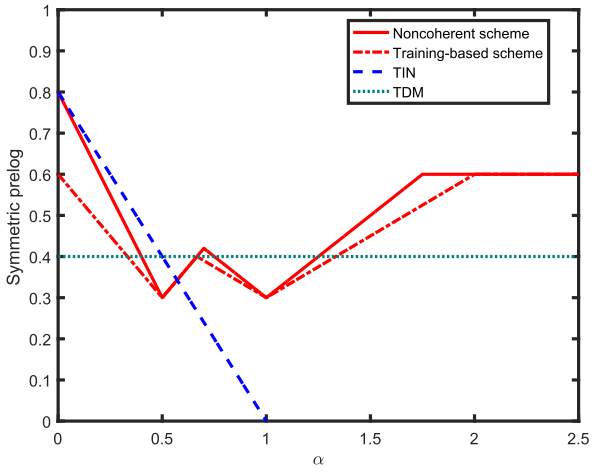

Fig. 8. Symmetric achievable prelog for coherence time $T = 5$.

TABLE III
COMPARISON OF RATES (BITS PER USER) ACHIEVABLE WITH DIFFERENT SCHEMES FOR $T = 5$, $\alpha = 1$

SNR dB	Rates for different schemes	
	2 symbol training	TDM
6	0.47	0.50
7	0.54	0.57
8	0.61	0.66
9	0.69	0.75
10	0.77	0.84

the expression in [24, (65)] for the point in the rate region where both users have the same rate. The expressions used for obtaining the rates for the TDM scheme are given in [24, Appendix E]. We also provide a Mathematica code at https://arxiv.org/src/1812.03579/anc/Noncoh_IC_rates.nb for calculating the rate points.

B. Coding Scheme and Analysis of the Rate Region

We describe the coding scheme starting with a general input distribution and then we evaluate the prelog region for Gaussian inputs.

1) Encoding: We consider a fixed distribution $p(\boldsymbol{U}_1)\,p(\boldsymbol{U}_2)\,p(\boldsymbol{X}_1|\boldsymbol{U}_1)\,p(\boldsymbol{X}_2|\boldsymbol{U}_2)$ where $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{X}_1, \boldsymbol{X}_2$ are vectors of length T. For Transmitter 1, generate $2^{NTR_{\mathrm{c1}}}$ codewords $\boldsymbol{U}_1^N(i)$ with $i \in \{1, \ldots, 2^{NTR_{\mathrm{c1}}}\}$ according to $\prod_{l=1}^{N} p\left(\boldsymbol{U}_{1(l)}\right)$. For each $\boldsymbol{U}_1^N(i)$, generate $2^{NTR_{\mathrm{p1}}}$ codewords $\boldsymbol{X}_1^N(i,j)$, with $j \in \{1, \ldots, 2^{NTR_{\mathrm{p1}}}\}$, according to $\prod_{l=1}^{N} p\left(\boldsymbol{X}_{1(l)}\middle|\boldsymbol{U}_{1(l)}\right)$. Similarly for Transmitter 2, generate $2^{NTR_{\mathrm{c2}}}$ codewords $\boldsymbol{U}_2^N(i)$, with $i \in \{1, \ldots, 2^{NTR_{\mathrm{c2}}}\}$, according to $\prod_{l=1}^{N} p\left(\boldsymbol{U}_{2(l)}\right)$. For each $\boldsymbol{U}_2^N(j)$, generate $2^{NTR_{\mathrm{p2}}}$ codewords $\boldsymbol{X}_2^N(i,j)$, with $j \in \{1, \ldots, 2^{NTR_{\mathrm{p2}}}\}$, according to $\prod_{l=1}^{N} p\left(\boldsymbol{X}_{2(l)}\middle|\boldsymbol{U}_{2(l)}\right)$.

Transmitter 1 has uniformly random messages $w_{\mathrm{c}1} \in \left\{1,\ldots,2^{NTR_{\mathrm{c}1}}\right\}, w_{\mathrm{p}1} \in \left\{1,\ldots,2^{NTR_{\mathrm{p}1}}\right\}$ to transmit and Transmitter 2 has uniformly random messages $w_{\mathrm{c}2} \in \left\{1,\ldots,2^{NTR_{\mathrm{c}2}}\right\}, w_{\mathrm{p}2} \in \left\{1,\ldots,2^{NTR_{\mathrm{p}2}}\right\}$ to transmit. Transmitter 1 sends the symbols $\boldsymbol{X}_1^N(w_{\mathrm{c}1}, w_{\mathrm{p}1})$ and Transmitter 2 sends the symbols $\boldsymbol{X}_1^N(w_{\mathrm{c}2}, w_{\mathrm{p}2})$.

2) Decoding: For decoding, Receiver 1 finds a triplet $\left(\hat{i},\hat{j},\hat{k}\right)$ requiring $\hat{i},\hat{j}$ to be unique with

$$\left(\boldsymbol{X}_1^N\left(\hat{i},\hat{j}\right), \boldsymbol{U}_1^N\left(\hat{i}\right), \boldsymbol{U}_2^N\left(\hat{k}\right), \boldsymbol{Y}_1^N\right) \in \mathcal{A}_\epsilon^{(N)}.$$

Similarly Receiver 2 finds a triplet $\left(\hat{i},\hat{j},\hat{k}\right)$ requiring $\hat{i},\hat{j}$ to be unique with

$$\left(\boldsymbol{X}_2^N\left(\hat{i},\hat{j}\right), \boldsymbol{U}_2^N\left(\hat{i}\right), \boldsymbol{U}_1^N\left(\hat{k}\right), \boldsymbol{Y}_2^N\right) \in \mathcal{A}_\epsilon^{(N)},$$

where $\mathcal{A}_\epsilon^{(N)}$ indicates the set of jointly typical sequences.

3) Error Analysis: We give the sketch of analysis for the error probability at Receiver 1 assuming $(i,j,k) = (1,1,1)$. Let $\mathcal{E}_{ijk}$ be the event $\left\{\left(\boldsymbol{X}_1^N(i,j), \boldsymbol{U}_1^N(i), \boldsymbol{U}_2^N(k), \boldsymbol{Y}_1^N\right) \in \mathcal{A}_\epsilon^{(N)}\right\}$ for a given i,j,k. By asymptotic equipartition property (AEP), the probability of $\Pr\left(\bigcup_k \mathcal{E}_{11k}\right)$ approaches unity. The error probability at Receiver 1 is then captured by the following:

$$\begin{aligned}
&\Pr\left(\bigcup_{(i,j)\neq(1,1)} \mathcal{E}_{ijk}\right)\\
&\le \left(\sum_{i\neq1,j\neq1,k\neq1} \Pr\left(\mathcal{E}_{ijk}\right) + \sum_{i\neq1,j=1,k\neq1} \Pr\left(\mathcal{E}_{ijk}\right)\right)\\
&\quad + \left(\sum_{i\neq1,j\neq1,k=1} \Pr\left(\mathcal{E}_{ijk}\right) + \sum_{i\neq1,j=1,k=1} \Pr\left(\mathcal{E}_{ijk}\right)\right)\\
&\quad + \sum_{i=1,j\neq1,k=1} \Pr\left(\mathcal{E}_{ijk}\right) + \sum_{i=1,j\neq1,k\neq1} \Pr\left(\mathcal{E}_{ijk}\right)\\
&\le 2^{N(TR_{\mathrm{c}1}+TR_{\mathrm{c}2}+TR_{\mathrm{p}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)}\\
&\quad + 2^{N(TR_{\mathrm{p}1}+TR_{\mathrm{c}1}-I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_2)+\epsilon)}\\
&\quad + 2^{N(TR_{\mathrm{p}1}-I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2)+\epsilon)}\\
&\quad + 2^{N(TR_{\mathrm{c}2}+TR_{\mathrm{p}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1)+\epsilon)}.
\end{aligned}$$

The details of the simplification in the last step can be obtained from standard textbooks, for example from [25, Ch. 6]. Requiring the average error probability to vanish at Receiver 1 and Receiver 2, we get the following equations as a sufficient condition:

$$TR_{\mathrm{c}1}+TR_{\mathrm{c}2}+TR_{\mathrm{p}1} \le I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1), \tag{11a}$$
$$TR_{\mathrm{p}1}+TR_{\mathrm{c}1} \le I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_2), \tag{11b}$$
$$TR_{\mathrm{p}1} \le I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2), \tag{11c}$$
$$TR_{\mathrm{c}2}+TR_{\mathrm{p}1} \le I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1), \tag{11d}$$
$$TR_{\mathrm{c}1}+TR_{\mathrm{c}2}+TR_{\mathrm{p}2} \le I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2), \tag{11e}$$
$$TR_{\mathrm{p}2}+TR_{\mathrm{c}2} \le I(\boldsymbol{X}_2;\boldsymbol{Y}_2|\boldsymbol{U}_1), \tag{11f}$$
$$TR_{\mathrm{p}2} \le I(\boldsymbol{X}_2;\boldsymbol{Y}_2|\boldsymbol{U}_1,\boldsymbol{U}_2), \tag{11g}$$
$$TR_{\mathrm{c}1}+TR_{\mathrm{p}2} \le I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2|\boldsymbol{U}_2). \tag{11h}$$

After Fourier-Motzkin elimination, the following equations are obtained for achievability, with $R_1 = R_{\mathrm{c}1}+R_{\mathrm{p}1}, R_2 = R_{\mathrm{c}2}+R_{\mathrm{p}2}$:

$$TR_1 \le I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_2), \tag{12a}$$
$$TR_2 \le I(\boldsymbol{X}_2;\boldsymbol{Y}_2|\boldsymbol{U}_1), \tag{12b}$$
$$T(R_1+R_2) \le I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2) + I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2), \tag{12c}$$
$$T(R_1+R_2) \le I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1) + I(\boldsymbol{X}_2;\boldsymbol{Y}_2|\boldsymbol{U}_1,\boldsymbol{U}_2), \tag{12d}$$
$$T(R_1+R_2) \le I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1) + I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2|\boldsymbol{U}_2), \tag{12e}$$
$$T(2R_1+R_2) \le I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1) + I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2) + I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2|\boldsymbol{U}_2), \tag{12f}$$
$$T(R_1+2R_2) \le I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2) + I(\boldsymbol{X}_2;\boldsymbol{Y}_2|\boldsymbol{U}_1,\boldsymbol{U}_2) + I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1). \tag{12g}$$

For power splitting, we adapt the idea of the simplified Han-Kobayashi scheme where the power allocation is such that the private signal is seen below the noise level at the other receiver. Similar to [16], [18], we choose $\boldsymbol{U}_k$ as a vector of length T with i.i.d. $\mathcal{CN}(0,\lambda_{\mathrm{c}})$ elements and $\boldsymbol{X}_{\mathrm{p}k}$ as a vector of length T with i.i.d. $\mathcal{CN}(0,\lambda_{\mathrm{p}})$ elements for $k\in\{1,2\}$. The random variables are chosen independent of each other so that the set $\{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{X}_{\mathrm{p}1},\boldsymbol{X}_{\mathrm{p}2}\}$ is mutually independent. We use $\boldsymbol{X}_1 = \boldsymbol{U}_1+\boldsymbol{X}_{\mathrm{p}1}$, $\boldsymbol{X}_2 = \boldsymbol{U}_2+\boldsymbol{X}_{\mathrm{p}2}$, $\lambda_{\mathrm{c}}+\lambda_{\mathrm{p}} = P$ and $\lambda_{\mathrm{p}} = P^{1-\alpha}$. The prelog results in Table I can be obtained by evaluating the rate region (12) for our choice of input distribution.

4) Preliminaries for Prelog Evaluation: We give some preliminary results that can be used in obtaining prelog region from our achievability region.

Fact 1: For an exponentially distributed random variable $\boldsymbol{\xi}$ with mean μ_ξ and with given constants $a\ge0, b>0$, we have

$$\log(a+b\mu_\xi) - \gamma\log(e) \le \mathbb{E}[\log(a+b\boldsymbol{\xi})] \tag{13}$$
$$\le \log(a+b\mu_\xi), \tag{14}$$

where γ is Euler's constant.

Proof: This is given in [18, Section III-B]. ∎

We now simplify the region (12) by considering the terms in it one by one.

Claim 1: The term $h(\boldsymbol{Y}_1|\boldsymbol{U}_2,\boldsymbol{X}_1)$ is upper bounded at high SNR *as*

$$h(\boldsymbol{Y}_1|\boldsymbol{U}_2,\boldsymbol{X}_1) \dot{\le} \log\left(\mathsf{SNR}+\mathsf{SNR}^\alpha\right) + \log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^\alpha\right)\right). \tag{15}$$

Proof: The outline of the proof is as follows: with $\boldsymbol{y}_{1,i}$ as the components of $\boldsymbol{Y}_1$, we expand $h(\boldsymbol{Y}_1|\boldsymbol{U}_2,\boldsymbol{X}_1) = \sum_i h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1,\{\boldsymbol{y}_{1,j}\}_{j=1}^i\right)$. The first term $h\left(\boldsymbol{y}_{1,1}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1\right)$ gives rise to the term $\log\left(\mathsf{SNR}+\mathsf{SNR}^\alpha\right)$ with uncertainty from both incoming channels.

Let us consider the term $h\left(\boldsymbol{y}_{1,2}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1,\boldsymbol{y}_{1,1}\right)$. In $\boldsymbol{y}_{1,2}$, the contribution to uncertainty is from the channels as well as from the symbols. When conditioned on $\boldsymbol{U}_2,\boldsymbol{X}_1$, the contribution of uncertainty from these symbols can be removed.

The uncertainty from $\boldsymbol{X}_{p2}$ in $\boldsymbol{y}_{1,2}$ can be neglected in prelog calculation due to the power allocation strategy that we use. The term $\boldsymbol{y}_{1,1}$ is a linear combination of the symbols as well as the channels. Using this single linear combination given in the conditioning, the uncertainty from one of the channels can be removed. Thus $h\left(\boldsymbol{y}_{1,2}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1,\boldsymbol{y}_{1,1}\right)$ gives rise to $\log\left(\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)=\log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right)$, with either the uncertainty from the direct channel removed or the uncertainty from the interfering channel removed.

In terms $h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1,\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}\right)$ with $i\geq 3$, we can follow the same procedure as stated in the above paragraph. However with $\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}$ available in the conditioning, we have more than a single linear combination of the channels available. Using these, the contribution from both channels can be removed, and hence $h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_2,\boldsymbol{X}_1,\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}\right)$ do not contribute to the prelog. The detailed proof is in [24, Appendix B]. ■

Claim 2: The term $h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)$ is lower bounded at high SNR *as*

$$\begin{aligned}h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)\dot{\geq}\ &\log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\&+\log\left(\mathsf{SNR}^{1-\alpha}+\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right)\\&+(T-2)\log\left(1+\mathsf{SNR}^{1-\alpha}\right).\end{aligned}$$

Proof: We expand $h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)=\sum_i h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_1,\boldsymbol{U}_2,\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}\right)$. One way to lower bound $h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_1,\boldsymbol{U}_2,\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}\right)$ is to condition on the channel strengths and reduce the term to that for a coherent channel. Another way to lower bound $h\left(\boldsymbol{y}_{1,i}\middle|\boldsymbol{U}_1,\boldsymbol{U}_2,\left\{\boldsymbol{y}_{1,j}\right\}_{j=1}^{i}\right)$ is to give all the transmit signals in the conditioning and reduce the entropy to that of a (conditionally) joint Gaussian distribution. These two techniques help us prove the claim. See Appendix I for more details. ■

5) Bounding Mutual Information Terms: In the following four claims, we obtain the lower bounds for four mutual information terms in the achievability region (12). We need to bound only four terms and the other terms can be bounded by using symmetry of the setup.

Claim 3: The term $I\left(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_2\right)$ is lower bounded at high SNR *as*

$$I\left(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_2\right)\dot{\geq}(T-1)\log\left(\mathsf{SNR}\right)-\log\left(\min(\mathsf{SNR},\mathsf{SNR}^{\alpha})\right).$$

Proof: We have

$$\begin{aligned}&h\left(\boldsymbol{Y}_1|\boldsymbol{U}_2\right)&(16)\\&=\ h\left(\boldsymbol{g}_{11}\boldsymbol{X}_1+\boldsymbol{g}_{21}\boldsymbol{X}_2+\boldsymbol{Z}_1|\boldsymbol{U}_2\right)&(17)\\&=\ \sum_{i=1}^{T}h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}\Big|\\&\qquad\left\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,j}+\boldsymbol{z}_{1,j}\right\}_{j=1}^{i-1},\boldsymbol{U}_2\Big)&(18)\\&\overset{(i)}{\geq}\ h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}|\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,1},\boldsymbol{U}_2\right)\\&\quad+\sum_{i=2}^{T}h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)&(19)\\&\overset{(ii)}{\geq}\ \log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)+(T-1)\log\left(\mathsf{SNR}\right),&(20)\end{aligned}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}\right)-\left(\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)-\left(\left\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,j}+\boldsymbol{z}_{1,j}\right\}_{j=1}^{i-1},\boldsymbol{U}_2\right)$. The step (ii) is using the property of Gaussians for the terms $h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}|\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,1},\boldsymbol{U}_2\right)$, $h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)=h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{\mathrm{p}2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)$ and using Fact 1. Using (20) and Claim 1 completes the proof. ■

Claim 4: The term $I\left(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2\right)$ is lower bounded at high SNR *as*

$$\begin{aligned}I\left(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2\right)\dot{\geq}\ &(T-1)\log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\&-\log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right).\end{aligned}$$

Proof: We have

$$h\left(\boldsymbol{Y}_2\right)\dot{\geq}T\log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right).\tag{21}$$

Using Claim 1 for $h\left(\boldsymbol{Y}_1|\boldsymbol{X}_1,\boldsymbol{U}_2\right)$ and using symmetry we get,

$$\begin{aligned}h\left(\boldsymbol{Y}_2|\boldsymbol{X}_2,\boldsymbol{U}_1\right)\dot{\leq}\ &\log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\&+\log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right).\end{aligned}\tag{22}$$

Combining the last two equations completes the proof. ■

Claim 5: The term $I\left(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)$ is lower bounded at high SNR *as*

$$\begin{aligned}I\left(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)\dot{\geq}\ &\log\left(\mathsf{SNR}^{1-\alpha}+\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right)\\&+\ (T-2)\log\left(1+\mathsf{SNR}^{1-\alpha}\right)\\&-\ \log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right).\end{aligned}$$

Proof: This follows by combining Claim 2 for $h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)$ and Claim 1 for $h\left(\boldsymbol{Y}_1|\boldsymbol{X}_1,\boldsymbol{U}_1,\boldsymbol{U}_2\right)$. ■

Claim 6: The term $I\left(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1\right)$ is lower bounded at high SNR *as*

$$\begin{aligned}I\left(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1|\boldsymbol{U}_1\right)\dot{\geq}\ &(T-1)\log\left(\mathsf{SNR}^{1-\alpha}+\mathsf{SNR}^{\alpha}\right)\\&-\log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right).\end{aligned}$$

Proof: We have

$$\begin{aligned}&h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1\right)\\&\quad=\ h\left(\boldsymbol{g}_{11}\boldsymbol{X}_1+\boldsymbol{g}_{21}\boldsymbol{X}_2+\boldsymbol{Z}_1|\boldsymbol{U}_1\right)\\&\quad=\ \sum_{i}h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}\Big|\\&\qquad\qquad\left\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,j}+\boldsymbol{z}_{1,j}\right\}_{j=1}^{i-1},\boldsymbol{U}_1\Big)\\&\quad\overset{(i)}{\geq}\ h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}|\boldsymbol{U}_1,\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,1}\right)\\&\qquad+\sum_{i=2}^{T}h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{u}_{1,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)\\&\quad\overset{(ii)}{\geq}\ \log\left(\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\&\qquad+\ (T-1)\log\left(\mathsf{SNR}^{1-\alpha}+\mathsf{SNR}^{\alpha}\right),\end{aligned}\tag{23}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}\right)-\left(\boldsymbol{u}_{1,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)-\left(\left\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,j}+\boldsymbol{z}_{1,j}\right\}_{j=1}^{i-1},\boldsymbol{U}_1\right)$. In step (ii) we

TABLE IV
LOWER BOUNDS AT HIGH SNR FOR THE TERMS IN THE ACHIEVABILITY REGION AND THEIR PRELOG

Term	Lower bound at high SNR	Prelog of lower bound		
		$\alpha < 1/2$	$1/2 < \alpha < 1$	$\alpha > 1$
$I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_2)$	$(T-1)\log(\mathsf{SNR}) - \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)-\alpha$	$(T-1)-\alpha$	$(T-2)$
$I(\boldsymbol{X}_2, \boldsymbol{U}_1; \boldsymbol{Y}_2)$	$(T-1)\log(\mathsf{SNR}+\mathsf{SNR}^{\alpha}) - \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)-\alpha$	$(T-1)-\alpha$	$(T-1)\alpha-1$
$I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2)$	$\log(\mathsf{SNR}^{1-\alpha}+\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha})) + (T-2)\log(1+\mathsf{SNR}^{1-\alpha}) - \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)(1-\alpha)-\alpha$	$(T-2)(1-\alpha)$	0
$I(\boldsymbol{X}_1, \boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1)$	$(T-1)\log(\mathsf{SNR}^{1-\alpha}+\mathsf{SNR}^{\alpha}) - \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)(1-\alpha)-\alpha$	$(T-2)\alpha$	$(T-1)\alpha-1$

TABLE V
ACHIEVABLE PRELOG REGION FOR DIFFERENT REGIMES OF α

$\alpha < 1/2$	$1/2 \leq \alpha \leq 1$	$\alpha > 1$
$d_1 \leq (1-1/T) - 2\alpha/T$ $d_2 \leq (1-1/T) - 2\alpha/T$ $d_1 + d_2 \leq 2(1-1/T) - \alpha(1+1/T)$	$d_1 \leq (1-2/T)$ $d_2 \leq (1-2/T)$ $d_1 + d_2 \leq (2-3/T) - \alpha(1-1/T)$	$d_1 + d_2 \leq (1-1/T)\alpha - 1/T$

removed the contribution of $\boldsymbol{g}_{11}\boldsymbol{u}_{1,i}$ from the second term and used the structure $\boldsymbol{x}_{1,i} = \boldsymbol{u}_{1,i} + \boldsymbol{x}_{1\mathrm{p},i}$, where $\boldsymbol{u}_{1,i}, \boldsymbol{x}_{1\mathrm{p},i}$ are independent Gaussian random variables and $\boldsymbol{x}_{1\mathrm{p},i}$ has variance $P^{1-\alpha} = \mathsf{SNR}^{1-\alpha}$. We also used Fact 1 together with the fact that the channels are Gaussian distributed.

We also have

$$\begin{aligned} h(\boldsymbol{Y}_1 \vert \boldsymbol{U}_2, \boldsymbol{U}_1, \boldsymbol{X}_1) &\leq h(\boldsymbol{Y}_1 \vert \boldsymbol{U}_2, \boldsymbol{X}_1) \\ &\dot{\leq} \log(\mathsf{SNR}+\mathsf{SNR}^{\alpha}) \\ &\quad + \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha})), \quad (24) \end{aligned}$$

using Claim 1 for $h(\boldsymbol{Y}_1 \vert \boldsymbol{X}_1, \boldsymbol{U}_2)$. Using (24) and (23) completes the proof. ∎

We collect the results from Claim 3, Claim 4, Claim 5 and Claim 6 in the second column of Table IV. In the third column of Table IV, we give the prelog for the lower bounds. Using the prelog of the lower bounds from Table IV in (12), using symmetry of the terms and using only the active inequalities, it can be verified that the prelog region in Table I is achievable.

III. NONCOHERENT IC WITH FEEDBACK

In this section, we provide our results for the noncoherent rate-splitting scheme for the noncoherent IC with feedback and compare the achievable prelog with a standard training-based scheme. We also compare the performance with the TIN and TDM schemes.

Theorem 3: For a noncoherent IC with feedback, the prelog region given in Table V is achievable:

Proof: This is obtained using the block Markov scheme of [17, Lemma 1] for the noncoherent case. We use a rate-splitting scheme based on the average interference-to-noise ratio and noncoherent decoding at the receivers. We use the block Markov scheme from [17, Lemma 1] with a total size of blocks B.

In block b, the message for User 1 is split into two parts, a common message $w_{\mathrm{c}1}^{(b)}$ at rate $R_{\mathrm{c}1}$ and a private message $w_{\mathrm{p}1}^{(b)}$ at rate $R_{\mathrm{p}1}$. The transmitted vector symbols at Transmitter 1 are of the form $\boldsymbol{X}_1 = \boldsymbol{U}_1 + \boldsymbol{X}_{\mathrm{p}1}$ where $\boldsymbol{U}_1, \boldsymbol{X}_{\mathrm{p}1}$ are independent Gaussian vectors of length T. The power of each element of $\boldsymbol{X}_{\mathrm{p}1}$ is $P^{1-\alpha}$ and the power of each element of $\boldsymbol{U}_1$ is $P - P^{1-\alpha}$. The Transmitter 1 is able to decode $w_{\mathrm{c}2}^{(b-1)}$ using feedback. The messages $w_{\mathrm{c}2}^{(b-1)}$, $w_{\mathrm{c}1}^{(b-1)}$ and $w_{\mathrm{c}1}^{(b)}$ are mapped into $\boldsymbol{U}_1$ in b^{th} block. The private message $w_{\mathrm{p}1}^{(b)}$ is mapped into $\boldsymbol{X}_{\mathrm{p}1}$.

For User 2, in block b, we have a common message $w_{\mathrm{c}2}^{(b)}$ at rate $R_{\mathrm{c}2}$ and a private message $w_{\mathrm{p}2}^{(b)}$ at rate $R_{\mathrm{p}2}$. The transmitted vector symbols at Transmitter 2 are of the form $\boldsymbol{X}_2 = \boldsymbol{U}_2 + \boldsymbol{X}_{\mathrm{p}2}$ where $\boldsymbol{U}_2, \boldsymbol{X}_{\mathrm{p}2}$ are independent Gaussian vectors of length T. The power of each element of $\boldsymbol{X}_{\mathrm{p}2}$ is $P^{1-\alpha}$ and the power of each element of $\boldsymbol{U}_2$ is $P - P^{1-\alpha}$. The Transmitter 2 is able to decode $w_{\mathrm{c}1}^{(b-1)}$ using feedback. The messages $w_{\mathrm{c}1}^{(b-1)}$, $w_{\mathrm{c}2}^{(b-1)}$ and $w_{\mathrm{c}2}^{(b)}$ are mapped into $\boldsymbol{U}_2$ in b^{th} block. The private message $w_{\mathrm{p}2}^{(b)}$ is mapped into $\boldsymbol{X}_{\mathrm{p}2}$.

The messages $w_{\mathrm{c}1}^{(b)}, w_{\mathrm{p}1}^{(b)}, w_{\mathrm{c}2}^{(b)}, w_{\mathrm{p}2}^{(b)}$ with $b = 0, B$ are set to be fixed and known to all transmitters and receivers. After B blocks, the receivers perform noncoherent backward decoding. Receiver 1 uses the symbols received in block b and decodes $w_{\mathrm{c}1}^{(b-1)}, w_{\mathrm{c}2}^{(b-1)}, w_{\mathrm{p}2}^{(b)}$ assuming $w_{\mathrm{c}1}^{(b)}, w_{\mathrm{c}2}^{(b)}$ are decoded from the symbols received in block $b+1$. Receiver 2 uses the symbols received in block b and decodes $w_{\mathrm{c}1}^{(b-1)}, w_{\mathrm{c}2}^{(b-1)}, w_{\mathrm{p}1}^{(b)}$ assuming $w_{\mathrm{c}1}^{(b)}, w_{\mathrm{c}2}^{(b)}$ are decoded from the symbols received in block $b+1$. The details of the coding scheme and its analysis are in Section III-B. ∎

We now obtain the prelog of a standard training-based scheme for the noncoherent IC with feedback.

Theorem 4: A standard training-based scheme for the noncoherent IC with feedback can achieve the prelog region described in Table VI.

Proof: For training, in every coherence period of T symbols, the first transmitter can send a known symbol while the second transmitter remains turned off; with this both

TABLE VI

ACHIEVABLE PRELOG REGION FOR DIFFERENT REGIMES OF α

$\alpha \leq 1$	$\alpha > 1$
$d_1 \leq (1-2/T)$ $d_2 \leq (1-2/T)$ $d_1 + d_2 \leq (1-2/T)(2-\alpha)$	$d_1 + d_2 \leq (1-2/T)\,\alpha$

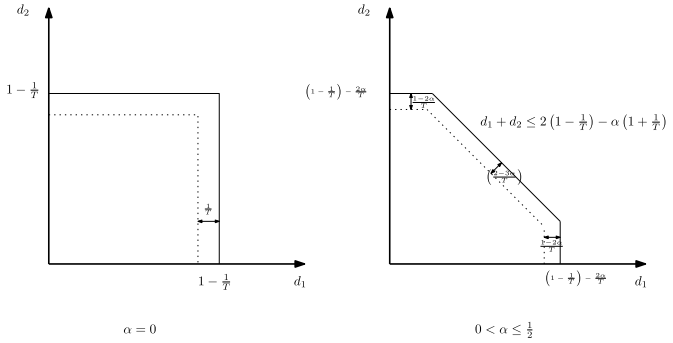

Fig. 9. Prelog region for $\alpha < 1/2, T \geq 2$. The solid line gives the prelog region achievable for a noncoherent scheme and the dotted line gives the prelog region for the scheme that uses 2 symbols for training.

receivers can estimate the channels from the first transmitter. Next the second transmitter can send a known symbol while the first transmitter remains turned off; with this both receivers can estimate the channels from the second transmitter. The rest of the symbols can be used to transmit data using a block Markov scheme similar to that described in Theorem 3. The detailed analysis for obtaining the prelog is given in [24, Appendix D]. ■

Remark 2: The capacity region of the coherent FF-IC with feedback is known within a constant gap from [18] and hence its gDoF region is known. The prelog region from the above theorem is the same as the gDoF region for the coherent case with a multiplication factor of $(1-2/T)$. *Hence the prelog obtained in Theorem 4 is the best among any scheme that uses two symbols for training in every coherence period of* T *symbols.*

A. Discussion

In Figure 9 and Figure 10, the prelog region achievable using our noncoherent scheme is compared with the prelog region achievable using the aforementioned training-based scheme. It can be observed that our noncoherent scheme outperforms the standard training-based scheme.

In Figure 11, we give the achievable symmetric prelog with coherence time $T = 3$ for our noncoherent rate-splitting scheme and the aforementioned training-based scheme for the feedback case. We give similar plots in Figure 12 for coherence time $T = 5$. We also include the prelog of the nonfeedback schemes from Section II in the figures. We had noticed in Section I that with feedback, the performance of noncoherent rate-splitting schemes is in general better than the TIN scheme. There are a few exceptions: when $T = 2$ and $\alpha < 1$, it can be calculated from Table V and [24, (71)] that the TIN scheme outperforms our noncoherent strategy with feedback. With $T = 3$ and $\alpha \leq .5$, our noncoherent rate-splitting strategy in the presence of feedback has the same prelog as the TIN scheme.

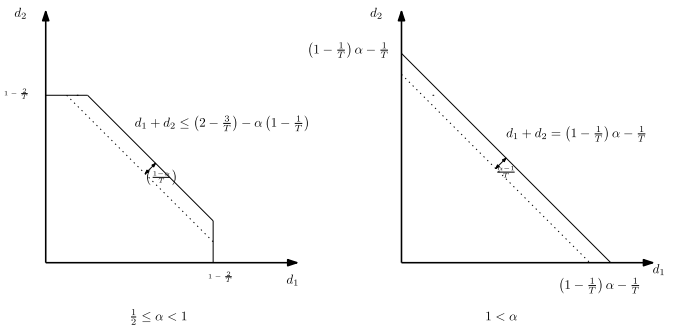

Fig. 10. Prelog region for $1/2 < \alpha$, $T \geq 2$. The solid line gives the prelog achievable for a noncoherent scheme and the dotted line is the prelog for a scheme that uses 2 symbols for training.

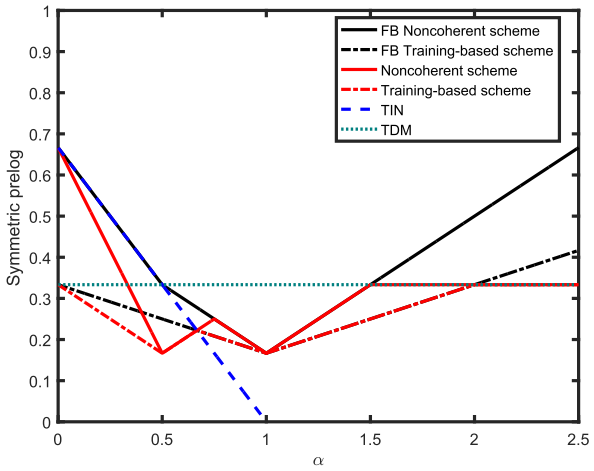

Fig. 11. Symmetric achievable prelog for coherence time $T = 3$: feedback and nonfeedback cases.

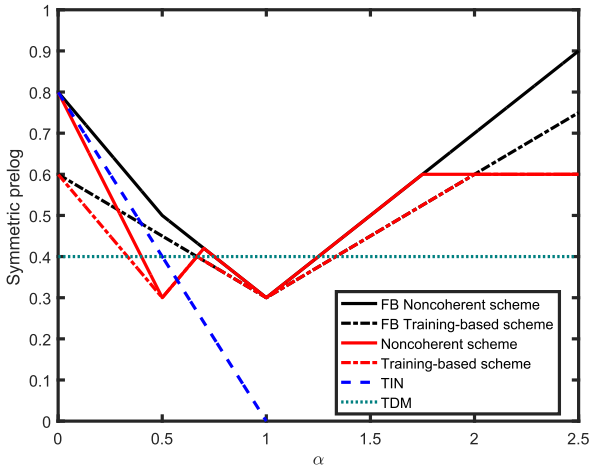

Fig. 12. Symmetric achievable prelog for coherence time $T = 5$: feedback and nonfeedback cases.

The noncoherent rate-splitting scheme attempts to decode part of the interfering message at the transmitter, and use it in subsequent transmissions. The rate that can be decoded at the transmitter using the feedback increases with T. For very weak interference level, the noncoherent rate-splitting scheme has a disadvantage as we explained in the discussion in Section I together with Figures 3 and 4. The advantage gained

by decoding at the transmitter outweighs this disadvantage when $T \geq 3$.

The TDM scheme outperforms other schemes for a region of α close to 1. This behavior can be explained similar to what we did in Section II-A. When $\alpha = 1$, the noncoherent rate-splitting scheme gives a prelog of $(1/2)(1-2/T)$ and the TDM scheme gives a prelog of $(1/2)(1-1/T)$. Hence for $\alpha = 1$, the noncoherent scheme effectively behaves as a TDM scheme that uses two symbols to train, but the TDM scheme can actually be implemented with only one training symbol.

B. Coding Scheme and Analysis of the Rate Region

We describe the coding scheme starting with a general input distribution and then we evaluate the prelog region for Gaussian inputs.

1) Encoding: Fix a joint distribution $p(\boldsymbol{U}_1)\,p(\boldsymbol{U}_2)\,p(\boldsymbol{X}_1|\boldsymbol{U}_1)\,p(\boldsymbol{X}_2|\boldsymbol{U}_2)$ where $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{X}_1, \boldsymbol{X}_2$ are vectors of length T. Generate $2^{NT(2R_{\mathrm{c}1}+R_{\mathrm{c}2})}$ codewords $\boldsymbol{U}_1^N(i,j,k)$ with $i,k \in \{1,\ldots,2^{NTR_{\mathrm{c}1}}\}$, $j \in \{1,\ldots,2^{NTR_{\mathrm{c}2}}\}$ according to $\prod_{l=1}^N p(\boldsymbol{U}_{1(l)})$. For each codeword $\boldsymbol{U}_1^N(i,j,k)$, generate $2^{NTR_{\mathrm{p}1}}$ codewords $\boldsymbol{X}_1^N(i,j,k,l)$ with $l \in \{1,\ldots,2^{NTR_{\mathrm{p}1}}\}$ according to $\prod_{l=1}^N p(\boldsymbol{X}_{1(l)}|\boldsymbol{U}_{1(l)})$.

Similarly generate $2^{NT(2R_{\mathrm{c}2}+R_{\mathrm{c}1})}$ codewords $\boldsymbol{U}_2^N(j,i,r)$ with $j,r \in \{1,\ldots,2^{NTR_{\mathrm{c}2}}\}$, $i \in \{1,\ldots,2^{NTR_{\mathrm{c}1}}\}$. For each codeword $\boldsymbol{U}_2^N(j,i,r)$, generate $2^{NTR_{\mathrm{p}2}}$ codewords $\boldsymbol{X}_2^N(j,i,r,s)$ with $s \in \{1,\ldots,2^{NTR_{\mathrm{p}2}}\}$ according to $\prod_{s=1}^N p(\boldsymbol{X}_{2(s)}|\boldsymbol{U}_{2(s)})$.

At block b, Transmitter 1 has uniformly random messages $w_{\mathrm{c}1}^{(b)} \in \{1,\ldots,2^{NTR_{\mathrm{c}1}}\}, w_{\mathrm{p}1}^{(b)} \in \{1,\ldots,2^{NTR_{\mathrm{p}1}}\}$ to transmit and Transmitter 2 has uniformly random messages $w_{\mathrm{c}2}^{(b)} \in \{1,\ldots,2^{NTR_{\mathrm{c}2}}\}, w_{\mathrm{p}2}^{(b)} \in \{1,\ldots,2^{NTR_{\mathrm{p}2}}\}$ to transmit. Using the symbols $\boldsymbol{Y}_1^{N,(b-1)}$ obtained through feedback, Transmitter 1 tries to noncoherently decode $\hat{w}_{2c}^{(b-1)} = \hat{k}$ by finding unique $\hat{k}$ such that

$$\Big(\boldsymbol{U}_1^N\big(w_{\mathrm{c}1}^{(b-2)}, w_{\mathrm{c}2}^{(b-2)}, w_{\mathrm{c}1}^{(b-1)}\big), \boldsymbol{X}_1^N\big(w_{\mathrm{c}1}^{(b-2)}, w_{\mathrm{c}2}^{(b-2)}, w_{\mathrm{c}1}^{(b-1)}, w_{\mathrm{p}1}^{(b-1)}\big), \boldsymbol{U}_2^N\big(w_{\mathrm{c}2}^{(b-2)}, w_{\mathrm{c}1}^{(b-2)}, \hat{k}\big), \boldsymbol{Y}_1^{N,(b-1)}\Big) \in \mathcal{A}_\epsilon^{(N)}.$$

where $\mathcal{A}_\epsilon^{(N)}$ indicates the set of jointly typical sequences. Transmitter 1 already knows $w_{\mathrm{c}1}^{(b-2)}, w_{\mathrm{c}1}^{(b-1)}, w_{\mathrm{p}1}^{(b-1)}$. Also $w_{\mathrm{c}2}^{(b-2)}$ is assumed to be correctly decoded in the previous block at Transmitter 1 and $w_{\mathrm{c}1}^{(b-2)}$ is assumed to be correctly decoded in the previous block at Transmitter 2. The current noncoherent decoding at Transmitter 1 is performed with vanishing error probability if

$$TR_{\mathrm{c}2} \leq I(\boldsymbol{U}_2; \boldsymbol{Y}_1|\boldsymbol{X}_1). \tag{25}$$

Based on $\hat{w}_{2c}^{(b-1)}$, Transmitter 1 then sends $\boldsymbol{X}_1^N\big(w_{\mathrm{c}1}^{(b-1)}, \hat{w}_{2c}^{(b-1)}, w_{\mathrm{c}1}^{(b)}, w_{\mathrm{p}1}^{(b)}\big)$. Similarly Transmitter 2 decodes $\hat{w}_{1c}^{(b-1)}$ and sends $\boldsymbol{X}_2^N\big(w_{\mathrm{c}2}^{(b-1)}, \hat{w}_{1c}^{(b-1)}, w_{\mathrm{c}2}^{(b)}, w_{\mathrm{p}2}^{(b)}\big)$. The messages $w_{\mathrm{c}1}^{(b)}, w_{\mathrm{p}1}^{(b)}, w_{\mathrm{c}2}^{(b)}, w_{\mathrm{p}2}^{(b)}$ for $b = 0, B$ can be set to be fixed and known to all transmitters and receivers.

2) Decoding: After receiving B blocks, each receiver performs backward decoding. At Receiver 1, block b is decoded assuming block $b+1$ is correctly decoded. From block $b+1$, $w_{\mathrm{c}2}^{(b)}, w_{1c}^{(b)}$ is assumed to be available at Receiver 1 after successful decoding. Now using the symbols from block b, Receiver 1 finds unique triplet $\left(\hat{i},\hat{j},\hat{l}\right)$ such that

$$\Big(\boldsymbol{U}_1^N\big(\hat{i},\hat{j},w_{\mathrm{c}1}^{(b)}\big), \boldsymbol{X}_1^N\big(\hat{i},\hat{j},w_{\mathrm{c}1}^{(b)},\hat{l}\big), \boldsymbol{U}_2^N\big(\hat{j},\hat{i},w_{\mathrm{c}2}^{(b)}\big), \boldsymbol{Y}_1^{N,(b)}\Big) \in \mathcal{A}_\epsilon^{(N)}.$$

Similarly Receiver 2 finds unique triplet $\left(\hat{j},\hat{i},\hat{s}\right)$ such that

$$\Big(\boldsymbol{U}_2^N\big(\hat{j},\hat{i},w_{\mathrm{c}1}^{(b)}\big), \boldsymbol{X}_2^N\big(\hat{j},\hat{i},w_{\mathrm{c}2}^{(b)},\hat{s}\big), \boldsymbol{U}_1^N\big(\hat{i},\hat{j},w_{\mathrm{c}1}^{(b)}\big), \boldsymbol{Y}_2^{N,(b)}\Big) \in \mathcal{A}_\epsilon^{(N)}.$$

3) Error Analysis: We give the sketch of error analysis at Receiver 1 assuming $\big(w_{\mathrm{c}1}^{(b-1)}, w_{\mathrm{c}2}^{(b-1)}, w_{\mathrm{p}1}^{(b)}\big) = (1,1,1)$ was sent through block $b-1$ and block b. We assume that there was no backward decoding error, *i.e.*, $\big(w_{\mathrm{c}1}^{(b)}, w_{\mathrm{c}2}^{(b)}\big)$ was correctly decoded. Let $\mathcal{E}_{ijl}$ be the event $\Big\{\Big(\boldsymbol{U}_1^N\big(\hat{i},\hat{j},w_{\mathrm{c}1}^{(b)}\big), \boldsymbol{X}_1^N\big(\hat{i},\hat{j},w_{\mathrm{c}1}^{(b)},\hat{k}\big), \boldsymbol{U}_2^N\big(\hat{j},\hat{i},w_{\mathrm{c}2}^{(b)}\big), \boldsymbol{Y}_1^{N,(b)}\Big) \in \mathcal{A}_\epsilon^{(N)}\Big\}$ for given i,j,l. By AEP, the probability of $\mathcal{E}_{111}$ approaches unity. The error probability is thus captured by the following equation using standard analysis similar to that in [17, Appendix B].

$$\begin{aligned}
&\Pr\Big(\bigcup_{(i,j,l)\neq(1,,1,1)} \mathcal{E}_{ijl}\Big) \\
&\leq \sum_{i\neq1,j\neq1,l\neq1} \Pr(\mathcal{E}_{ijl}) + \sum_{i=1,j=1,l\neq1} \Pr(\mathcal{E}_{ijl}) \\
&\quad + \sum_{i=1,j\neq1,l=1} \Pr(\mathcal{E}_{ijl}) + \sum_{i\neq1,j=1,l=1} \Pr(\mathcal{E}_{ijl}) \\
&\quad + \sum_{i\neq1,j\neq1,l=1} \Pr(\mathcal{E}_{ijl}) + \sum_{i\neq1,j=1,l\neq1} \Pr(\mathcal{E}_{ijl}) \\
&\quad + \sum_{i=1,j\neq1,l\neq1} \Pr(\mathcal{E}_{ijl}) \\
&\leq 2^{N(TR_{\mathrm{c}1}+TR_{\mathrm{c}2}+TR_{\mathrm{p}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)} \\
&\quad + 2^{N(TR_{\mathrm{p}1}-I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2)+\epsilon)} \\
&\quad + 2^{N(TR_{\mathrm{c}2}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)} + 2^{N(TR_{\mathrm{c}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)} \\
&\quad + 2^{N(TR_{\mathrm{c}1}+TR_{\mathrm{c}2}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)} \\
&\quad + 2^{N(TR_{\mathrm{c}1}+TR_{\mathrm{p}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)} \\
&\quad + 2^{N(TR_{\mathrm{c}2}+TR_{\mathrm{p}1}-I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1)+\epsilon)}
\end{aligned} \tag{26}$$

Combining (25) and (26), and considering similar analysis for User 2, we get the following equations for achievability:

$$TR_{\mathrm{c}2} \leq I(\boldsymbol{U}_2; \boldsymbol{Y}_1|\boldsymbol{X}_1), \tag{27a}$$
$$TR_{\mathrm{p}1} \leq I(\boldsymbol{X}_1; \boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2), \tag{27b}$$
$$T(R_{\mathrm{c}1}+R_{\mathrm{c}2}+R_{\mathrm{p}1}) \leq I(\boldsymbol{X}_1,\boldsymbol{U}_2;\boldsymbol{Y}_1), \tag{27c}$$
$$TR_{\mathrm{c}1} \leq I(\boldsymbol{U}_1; \boldsymbol{Y}_2|\boldsymbol{X}_2), \tag{27d}$$
$$TR_{\mathrm{p}2} \leq I(\boldsymbol{X}_2; \boldsymbol{Y}_2|\boldsymbol{U}_2,\boldsymbol{U}_1), \tag{27e}$$
$$T(R_{\mathrm{c}1}+R_{\mathrm{c}2}+R_{\mathrm{p}2}) \leq I(\boldsymbol{X}_2,\boldsymbol{U}_1;\boldsymbol{Y}_2). \tag{27f}$$

TABLE VII
LOWER BOUNDS AT HIGH SNR FOR THE TERMS IN THE ACHIEVABILITY REGION AND THEIR PRELOG

Term	Lower bound at high SNR	Prelog of lower bound		
		$\alpha < 1/2$	$1/2 < \alpha < 1$	$\alpha > 1$
$I(\boldsymbol{X}_2, \boldsymbol{U}_1; \boldsymbol{Y}_2)$	$(T-1)\log(\mathsf{SNR}+\mathsf{SNR}^{\alpha})$ $-\log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)-\alpha$	$(T-1)-\alpha$	$(T-1)\alpha - 1$
$I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2)$	$\log\left(\mathsf{SNR}^{1-\alpha}+\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha})\right)$ $+(T-2)\log\left(1+\mathsf{SNR}^{1-\alpha}\right)$ $-\log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-1)(1-\alpha)-\alpha$	$(T-2)(1-\alpha)$	0
$I(\boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{X}_1)$	$(T-1)\log(\mathsf{SNR}^{\alpha})$ $-\log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$	$(T-2)\alpha$	$(T-2)\alpha$	$(T-1)\alpha - 1$

After performing Fourier-Motzkin elimination similar to that in [17, Appendix B], we obtain the following achievability region with $R_1 = R_{\mathrm{c}1} + R_{\mathrm{p}1}, R_2 = R_{\mathrm{c}2} + R_{\mathrm{p}2}$:

$$TR_1 \le I(\boldsymbol{X}_1, \boldsymbol{U}_2; \boldsymbol{Y}_1), \tag{28a}$$

$$TR_1 \le I(\boldsymbol{U}_1; \boldsymbol{Y}_2 \vert \boldsymbol{X}_2) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2), \tag{28b}$$

$$TR_2 \le I(\boldsymbol{X}_2, \boldsymbol{U}_1; \boldsymbol{Y}_2), \tag{28c}$$

$$TR_2 \le I(\boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{X}_1) + I(\boldsymbol{X}_2; \boldsymbol{Y}_2 \vert \boldsymbol{U}_1, \boldsymbol{U}_2), \tag{28d}$$

$$T(R_1+R_2) \le I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2) + I(\boldsymbol{X}_2, \boldsymbol{U}_1; \boldsymbol{Y}_2), \tag{28e}$$

$$T(R_1+R_2) \le I(\boldsymbol{X}_2; \boldsymbol{Y}_2 \vert \boldsymbol{U}_1, \boldsymbol{U}_2) + I(\boldsymbol{X}_1, \boldsymbol{U}_2; \boldsymbol{Y}_1). \tag{28f}$$

We choose $\boldsymbol{U}_k$ as a vector of length T with i.i.d. $\mathcal{CN}(0, \lambda_{\mathrm{c}})$ elements and $\boldsymbol{X}_{\mathrm{p}k}$ as a vector of length T with i.i.d. $\mathcal{CN}(0, \lambda_{\mathrm{p}})$ elements for $k \in \{1, 2\}$. The random variables are chosen independent of each other so that the set $\{\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{X}_{\mathrm{p}1}, \boldsymbol{X}_{\mathrm{p}2}\}$ is mutually independent. We use $\boldsymbol{X}_1 = \boldsymbol{U}_1 + \boldsymbol{X}_{\mathrm{p}1}$, $\boldsymbol{X}_2 = \boldsymbol{U}_2 + \boldsymbol{X}_{\mathrm{p}2}$ where $\lambda_{\mathrm{c}} + \lambda_{\mathrm{p}} = P$ and $\lambda_{\mathrm{p}} = \min(P/\mathsf{INR}, P) = \min\left(P^{1-\alpha}, P\right)$ similar to [17], [18]. For prelog characterization, we can assume $P^{\alpha} > 1$. Hence, we have $\lambda_{\mathrm{p}} = P^{1-\alpha}$.

The prelog results in Table V can be obtained by evaluating the rate region (28) for our choice of input distribution. Note that the joint distribution of $(\boldsymbol{X}_1, \boldsymbol{Y}_1, \boldsymbol{U}_1, \boldsymbol{X}_2, \boldsymbol{Y}_2, \boldsymbol{U}_2)$ in its *single letter form* is the same as that for the nonfeedback case in Section II-B, hence we can carry over the inequalities for the single letter mutual information terms from Section II-B. We will use Claim 4 and Claim 5 from Section II-B to bound $I(\boldsymbol{X}_2, \boldsymbol{U}_1; \boldsymbol{Y}_2)$ and $I(\boldsymbol{X}_1; \boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2)$ respectively. We bound the term $I(\boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{X}_1)$ with the following claim.

Claim 7: The term $I(\boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{X}_1)$ is lower bounded at high SNR *as*

$$I(\boldsymbol{U}_2; \boldsymbol{Y}_1 \vert \boldsymbol{X}_1) \dot{\ge} (T-1)\log(\mathsf{SNR}^{\alpha}) - \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha})).$$

Proof: We have

$$\begin{aligned} h(\boldsymbol{Y}_1 \vert \boldsymbol{X}_1) &= h(\boldsymbol{g}_{11}\boldsymbol{X}_1 + \boldsymbol{g}_{21}\boldsymbol{X}_2 + \boldsymbol{Z}_1 \vert \boldsymbol{X}_1) \\ &= \sum_i h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,i} + \boldsymbol{z}_{1,i} \Big\vert \{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,j} + \boldsymbol{z}_{1,j}\}_{j=1}^{i-1}, \boldsymbol{X}_1\Big) \\ &\overset{(i)}{\dot{\ge}} h(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,1} + \boldsymbol{z}_{1,1} \vert \boldsymbol{x}_{2,1}, \boldsymbol{X}_1) + \sum_{i=2}^{T} h(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,i} + \boldsymbol{z}_{1,i} \vert \boldsymbol{X}_1, \boldsymbol{g}_{21}, \boldsymbol{g}_{11}) \\ &\overset{(ii)}{\dot{\ge}} \log(\mathsf{SNR}+\mathsf{SNR}^{\alpha}) + (T-1)\log(\mathsf{SNR}^{\alpha}), \end{aligned} \tag{29}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,i} + \boldsymbol{z}_{1,i}) - (\boldsymbol{X}_1, \boldsymbol{g}_{21}, \boldsymbol{g}_{11}) - \left(\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,j} + \boldsymbol{z}_{1,j}\}_{j=1}^{i-1}, \boldsymbol{X}_1\right)$ and (ii) is using the property of Gaussians for the terms $h(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,1} + \boldsymbol{z}_{1,1} \vert \boldsymbol{x}_{2,1}, \boldsymbol{X}_1)$, $h(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,i} + \boldsymbol{z}_{1,i} \vert \boldsymbol{X}_1, \boldsymbol{g}_{21}, \boldsymbol{g}_{11})$ and using Fact 1. Using (29) and $h(\boldsymbol{Y}_1 \vert \boldsymbol{U}_2, \boldsymbol{X}_1) \dot{\le} \log(\mathsf{SNR}+\mathsf{SNR}^{\alpha}) + \log(\min(\mathsf{SNR}, \mathsf{SNR}^{\alpha}))$ from Claim 1 completes the proof. ∎

Using Claim 4, Claim 5 and Claim 7, we have the lower bounds for the terms in the achievability region in the second column of Table VII. In the third column of Table VII, we obtain the prelog for the lower bounds. Using the prelog of the lower bounds from Table VII in (28), using symmetry of the terms and using only the active inequalities, it can be verified that the prelog region in Table V is achievable.

IV. CONCLUSION AND REMARKS

We studied the 2-user noncoherent IC with different achievability strategies. We observed that a standard training-based scheme is suboptimal in terms of prelog. Depending on the level of interference, a noncoherent scheme or a TIN scheme or a TDM scheme can give superior performance than the standard training-based scheme. Thus, the result for single user noncoherent channels that training-based schemes are DoF optimal does not extend to the gDoF of the noncoherent IC. Our current results are on inner bounds, outer bounds are still open.

APPENDIX I
PROOF OF CLAIM 2

In this appendix, we obtain a high-SNR bound on $h(\boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2)$. We have

$$\begin{aligned} h(\boldsymbol{Y}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2) &= h(\boldsymbol{g}_{11}\boldsymbol{X}_1 + \boldsymbol{g}_{21}\boldsymbol{X}_2 + \boldsymbol{Z}_1 \vert \boldsymbol{U}_1, \boldsymbol{U}_2) \\ &= \sum_i h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,i} + \boldsymbol{z}_{1,i} \Big\vert \{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j} + \boldsymbol{g}_{21}\boldsymbol{x}_{2,j} + \boldsymbol{z}_{1,j}\}_{j=1}^{i-1}, \boldsymbol{U}_1, \boldsymbol{U}_2\Big) \end{aligned} \tag{30}$$

$$
\begin{aligned}
&\overset{(i)}{\geq} h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}|\boldsymbol{x}_{1,1},\boldsymbol{x}_{2,1},\boldsymbol{U}_1,\boldsymbol{U}_2\right)\\
&\quad+h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\Big|\\
&\qquad\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2\Big)\\
&\quad+\sum_{i=3}^{T}h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{u}_{1,i},\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)\\
&\overset{(ii)}{\geq} \log\left(1+\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\
&\quad+h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\\
&\qquad\Big|\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2\Big)\\
&\quad+(T-2)\log\left(1+\mathsf{SNR}^{1-\alpha}\right),
\end{aligned}
\tag{31}
$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i})-(\boldsymbol{u}_{1,i},\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11})-\left(\{\boldsymbol{g}_{11}\boldsymbol{x}_{1,j}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,j}+\boldsymbol{z}_{1,j}\}_{j=1}^{i-1},\boldsymbol{U}_1,\boldsymbol{U}_2\right)$ and (ii) is using the property of Gaussians and using Fact 1. In (ii) for the last term, we use

$$
\begin{aligned}
&h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{u}_{1,i},\boldsymbol{u}_{2,i},\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)\\
&\overset{(i)}{=} h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{\mathrm{p}1,i}+\boldsymbol{g}_{21}\boldsymbol{x}_{\mathrm{p}2,i}+\boldsymbol{z}_{1,i}|\boldsymbol{g}_{21},\boldsymbol{g}_{11}\right)\\
&=\mathbb{E}\left[\log\left(\pi e\left(1+P^{1-\alpha}\left|\boldsymbol{g}_{11}\right|^2+P^{1-\alpha}\left|\boldsymbol{g}_{21}\right|^2\right)\right)\right]\\
&\overset{(ii)}{\doteq}\log\left(1+P^{1-\alpha}\right)=\log\left(1+\mathsf{SNR}^{1-\alpha}\right).
\end{aligned}
\tag{32}
\tag{33}
$$

(i) is by removing $\boldsymbol{g}_{11}\boldsymbol{u}_{1,i}+\boldsymbol{g}_{21}\boldsymbol{u}_{2,i}$ that is available in the conditioning and because the private message parts $\boldsymbol{x}_{\mathrm{p}1,i}$, $\boldsymbol{x}_{\mathrm{p}2,i}$ are independent of the common message parts $\boldsymbol{u}_{1,i},\boldsymbol{u}_{2,i}$. The step (ii) is using Fact 1. Now,

$$
\begin{aligned}
&h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\Big|\\
&\qquad\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2\Big)
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
&\geq h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\Big|\\
&\qquad\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{U}_1,\boldsymbol{U}_2\Big)\\
&=h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2},\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}\Big|\\
&\qquad\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{U}_1,\boldsymbol{U}_2\Big)\\
&\quad-h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1}\Big|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{U}_1,\boldsymbol{U}_2\Big)
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
&\overset{(i)}{=}\mathbb{E}\left[\log\left(\pi e\left|\begin{array}{c}\left|\boldsymbol{x}_{1,2}\right|^2+P^{\alpha-1}\left|\boldsymbol{x}_{2,2}\right|^2+1\\ \left(\boldsymbol{x}_{1,2}\boldsymbol{x}_{1,1}^{\dagger}+P^{\alpha-1}\boldsymbol{x}_{2,2}\boldsymbol{x}_{2,1}^{\dagger}\right)^{\dagger}\end{array}\right.\right.\right.\\
&\qquad\left.\left.\left.\begin{array}{c}\boldsymbol{x}_{1,2}\boldsymbol{x}_{1,1}^{\dagger}+P^{\alpha-1}\boldsymbol{x}_{2,2}\boldsymbol{x}_{2,1}^{\dagger}\\ \left|\boldsymbol{x}_{1,1}\right|^2+P^{\alpha-1}\left|\boldsymbol{x}_{2,1}\right|^2+1\end{array}\right|\right)\right]\\
&\quad-\mathbb{E}\left[\log\left(1+\left|\boldsymbol{x}_{2,1}\right|^2P^{\alpha-1}+\left|\boldsymbol{x}_{1,1}\right|^2\right)\right]
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
&\geq\mathbb{E}\Big[\log\Big(P^{\alpha-1}\Big(\left|\boldsymbol{x}_{1,1}\right|^2\left|\boldsymbol{x}_{2,2}\right|^2+\left|\boldsymbol{x}_{1,2}\right|^2\left|\boldsymbol{x}_{2,1}\right|^2\\
&\qquad-2\mathrm{Re}\left(\boldsymbol{x}_{1,2}\boldsymbol{x}_{1,1}^{\dagger}\boldsymbol{x}_{2,2}^{\dagger}\boldsymbol{x}_{2,1}\right)\Big)\Big)\Big]\\
&\quad-\log\left(1+P\cdot P^{\alpha-1}+P\right)
\end{aligned}
$$

$$
\begin{aligned}
&=\log\left(P^{\alpha-1}\right)+\mathbb{E}\left[\log\left(\left|\boldsymbol{x}_{1,1}\boldsymbol{x}_{2,2}-\boldsymbol{x}_{1,2}\boldsymbol{x}_{2,1}\right|^2\right)\right]\\
&\quad-\log\left(1+P^{\alpha}+P\right)
\end{aligned}
\tag{37}
$$

$$
\overset{(ii)}{\doteq}\log\left(P^{\alpha-1}\right)+\log\left(P^2\right)-\log\left(1+P^{\alpha}+P\right)
\tag{38}
$$

$$
\overset{(iii)}{\doteq}\log\left(\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right),
\tag{39}
$$

where (i) is using the property of Gaussian random variables, (ii) is using Fact 1 on page 8 and Tower property of expectation for $\mathbb{E}\left[\log\left(\left|\boldsymbol{x}_{1,1}\boldsymbol{x}_{2,2}-\boldsymbol{x}_{1,2}\boldsymbol{x}_{2,1}\right|^2\right)\right]$, (iii) is using our system setting with $P=\mathsf{SNR}$. Also

$$
\begin{aligned}
&h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\Big|\\
&\qquad\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2\Big)
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
&\overset{(i)}{\geq} h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\\
&\qquad\Big|\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{g}_{11},\boldsymbol{g}_{21}\Big)
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
&\overset{(ii)}{=} h\left(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}|\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{g}_{11},\boldsymbol{g}_{21}\right)\\
&\overset{(ii)}{\geq}\log\left(1+\mathsf{SNR}^{1-\alpha}\right),
\end{aligned}
\tag{42}
$$

where (i) is using the fact that conditioning reduces entropy, (ii) is due to the Markov chain $(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2})-(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{g}_{21},\boldsymbol{g}_{11})-(\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2)$, (iii) is following similar steps as for (33). Now combining (42), (39), we get

$$
\begin{aligned}
&h\Big(\boldsymbol{g}_{11}\boldsymbol{x}_{1,2}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,2}+\boldsymbol{z}_{1,2}\Big|\\
&\qquad\boldsymbol{g}_{11}\boldsymbol{x}_{1,1}+\boldsymbol{g}_{21}\boldsymbol{x}_{2,1}+\boldsymbol{z}_{1,1},\boldsymbol{U}_1,\boldsymbol{U}_2\Big)\\
&\dot{\geq}\log\left(1+\mathsf{SNR}^{1-\alpha}+\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right).
\end{aligned}
$$

Hence, substituting the above equation in (31), we get

$$
\begin{aligned}
&h\left(\boldsymbol{Y}_1|\boldsymbol{U}_1,\boldsymbol{U}_2\right)\\
&\quad\dot{\geq}\log\left(1+\mathsf{SNR}+\mathsf{SNR}^{\alpha}\right)\\
&\qquad+\log\left(1+\mathsf{SNR}^{1-\alpha}+\min\left(\mathsf{SNR},\mathsf{SNR}^{\alpha}\right)\right)\\
&\qquad+(T-2)\log\left(1+\mathsf{SNR}^{1-\alpha}\right).
\end{aligned}
$$

REFERENCES

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[2] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[3] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[4] T. Koch and G. Kramer, "On noncoherent fading relay channels at high signal-to-noise ratio," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2221–2241, Apr. 2013.

[5] J. Sebastian and S. Diggavi, "Generalized degrees of freedom of noncoherent diamond networks," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 5228–5260, Aug. 2020.

[6] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, Jul. 1997.

[7] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[8] R. H. Gohary and H. Yanikomeroglu, "Grassmannian signalling achieves tight bounds on the ergodic high-SNR capacity of the noncoherent MIMO full-duplex relay channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2480–2494, May 2014.
[9] J. Sebastian, A. Sengupta, and S. N. Diggavi, "On capacity of noncoherent MIMO with asymmetric link strengths," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 541–545.
[10] J. Sebastian and S. N. Diggavi, "Generalized degrees freedom of noncoherent MIMO channels with asymmetric link strengths," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4431–4448, Jul. 2020.
[11] A. Ozgur and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 669–673.
[12] A. Ozgur and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice-based QMF codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8275–8294, Dec. 2013.
[13] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[14] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.
[15] H.-F. Chong, M. Motani, H. K. Garg, and H. E. Gamal, "On the Han–Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3194, Jun. 2008.
[16] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
[17] C. Suh and D. N. C. Tse, "Feedback capacity of the Gaussian interference channel to within 2 bits," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2667–2685, May 2011.
[18] J. Sebastian, C. Karakus, and S. Diggavi, "Approximate capacity of fast fading interference channels with no instantaneous CSIT," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6015–6027, Dec. 2018.
[19] A. Vahid, C. Suh, and A. S. Avestimehr, "Interference channels with rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2788–2812, May 2012.
[20] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, Feb. 2009.
[21] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.
[22] S. Karmakar and M. K. Varanasi, "The generalized degrees of freedom region of the MIMO interference channel and its achievability," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7188–7203, Dec. 2012.
[23] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[24] J. Sebastian and S. Diggavi, "On the generalized degrees of freedom of noncoherent interference channel," 2018, *arXiv:1812.03579*.
[25] A. Gamal and Y. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011. [Online]. Available: http://books.google.com/books?id=l31D4DU7jykC

Joyson Sebastian received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 2012, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Los Angeles (UCLA), USA, in 2015 and 2018, respectively. He is currently an Engineer with Samsung. His research interests include information theory, wireless networks, and algorithms. He was a recipient of UCLA Graduate Division Fellowship in 2013 and the Guru Krupa Fellowship in 2014 and 2018.

Suhas Diggavi (Fellow, IEEE) received the degree from IIT Delhi and the Ph.D. degree from Stanford University.

He has worked as a Principal Member Research Staff at AT&T Shannon Laboratories and directed the Laboratory for Information and Communication Systems (LICOS), EPFL. He is currently a Professor of electrical and computer engineering at UCLA, where he directs the Information Theory and Systems Laboratory. He has eight issued patents. His research interests include information theory and its applications to several areas, including machine learning, security and privacy, wireless networks, data compression, cyber-physical systems, and bio-informatics and neuroscience (more information can be found at http://licos.ee.ucla.edu).

Dr. Diggavi has received several recognitions for his research from IEEE and ACM, including the 2013 IEEE Information Theory Society & Communications Society Joint Paper Award, the 2021 ACM Conference on Computer and Communications Security (CCS) Best Paper Award, the 2013 ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) Best Paper Award, and the 2006 IEEE Donald Fink Prize Paper Award among others. He was selected as a Guggenheim Fellow in 2021. He also received the 2019 Google Faculty Research Award, the 2020 Amazon Faculty Research Award, and the 2021 Facebook Faculty Research Award. He served as an IEEE Distinguished Lecturer and also currently serves on board of governors for the IEEE Information Theory Society. He has also helped organize IEEE and ACM conferences, including serving as the Technical Program Co-Chair for the 2012 IEEE Information Theory Workshop (ITW), the Technical Program Co-Chair for the 2015 IEEE International Symposium on Information Theory (ISIT), and General Co-Chair for the Mobihoc 2018. He has been an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE/ACM TRANSACTIONS ON NETWORKING, and other journals and special issues, as well as in the program committees of several IEEE conferences.