IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 8, 2019, accepted August 16, 2019, date of publication August 21, 2019, date of current version September 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936408

Interference Channel With Full-Duplex Amplify-and-Forward Transmitter Cooperations

JIANHAO HUANG[1], DAN WANG[2], AND CHUAN HUANG[1], (Member, IEEE)
[1]School of Electrical Engineering and Intelligentization, Dongguan University of Technology, Dongguan 523808, China
[2]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Chuan Huang (huangch@uestc.edu.cn)

ABSTRACT This paper considers a Gaussian interference channel (IC), where two transmitters send two messages to two receivers respectively. By utilizing the in-band full-duplex (FD) amplify-and-forward (AF) scheme, transmitter cooperations are deployed to enhance the system performance. In particular, each transmitter is capable of listening to its counterpart and then simultaneously transmits both the source signals to achieve higher power gain. Based on the proposed scheme, the equivalent channel model as well as the statistics of the accumulated residual self-interference and noise (ARIN) introduced by the transmitter cooperations and imperfect self-interference (SI) cancellation is analyzed. With the joint and single-user decoding schemes, the corresponding achievable rate regions are derived and a two-stage iterative algorithm is proposed to characterize these regions: in each step, fix the covariance matrix of the interferences and noises, and then adopt a two-step iterative semidefinite relaxation (SDR) method to optimize the two transmitters' transmission parameters, respectively. Simulation results reveal that the proposed scheme can significantly improve the achievable rate under several channel conditions.

INDEX TERMS Interference channel, full-duplex (FD), amplify-and-forward (AF), rate region, semidefinite relaxation (SDR).

I. INTRODUCTION

For the forthcoming 5G cellular systems, interference management is one of the most crucial tasks for physical-layer transmissions and network-level user scheduling, especially in the ultra dense networks, where numerous randomly deployed access points and mobile users are located within an area of interest [1]. As the network becomes denser, interferences between the access points and mobile users become more complicated. For example, when two mobile users are closely located at the edge of two cells, respectively, co-channel interference, which is very strong due to the similar path losses between the access points and the mobile users, becomes a major obstacle in increasing the network throughput.

From the information theoretical view point, multiple transmission and reception pairs sharing the same frequency band and interfering with each other is modeled as the interference channel (IC); this channel was initially studied by Shannon [2] and later by Carleial [3]. For the two-user IC case, the best known achievable result was obtained by Han and Kobayashi by the following scheme [4]: Each source message is split into two sub-messages, one as a common message decoded by both the receivers and one as a private message decoded by the intended receiver; each receiver decodes its desired message and the common message from the other transmitter, and the private message from the undesired receiver is treated as interference. In particular, when the co-channel interference is very weak [5], single-user decoding scheme, for which each receiver only decodes its own desired message and simply treats the other transmitter's signal as interference, is optimal to achieve the channel capacity [5], [6]; on the other hand, when the co-channel interference is very strong [7], joint decoding of both the source messages at each receiver is optimal to achieve the channel capacity [8].

Furthermore, literature [9]–[15] showed that by introducing transmitter cooperations, the achievable rates of the IC can be potentially improved. Transmitter cooperations can

The associate editor coordinating the review of this article and approving it for publication was Prabhat Kumar Upadhyay.

be established in two different manners: out-of-band conferencing links [9], [11], [12] and in-band full-duplex (FD) communications [13]–[15]. For the first case, transmitters talk to each other via conferencing links within limited communication rates, and the transmitter cooperation links are orthogonal to the transmitter-receiver channels. The authors in [9] extended the conferencing strategy proposed by Willems in [16] to the case of the Gaussian compound multiple access channel (MAC) and derived the corresponding capacity region. Wang and Tse [11] investigated the benefits of limited transmitter cooperations for the Gaussian IC by further dividing messages into noncooperative and cooperative ones, and applying linear beamforming for the cooperative messages, and the gap between the achievable rate and channel capacity was shown to be bounded with finite bits. For the symmetric Gaussian IC, Bagheri *et al.* [12] proposed a transmission strategy following the Han-Kobayashi scheme, while applying zero-forcing to partially suppress the private messages at each receiver. It is worth pointing out that the aforementioned transmitter cooperations are based on the decode-and-forward (DF) scheme [17], [18] and each transmitter (partially) decodes the messages from its counterpart via the conferencing links.

In contrast to the aforementioned case, transmitter cooperations utilizing the in-band FD communications, where the transmitters transmit and receive simultaneously at the same frequency band, will not consume extra frequency resources, while it may introduce more interference. For the two-user Gaussian IC with source cooperations, Høst-Madsen [13] derived the upper bounds for the capacity region and the DF scheme and dirty paper coding were applied to obtain the achievable rate for both the asynchronous and synchronous transmitter cases, respectively. Under the same IC setup, Prabhakaran and Viswanath [14] proposed a coding scheme based on the Han-Kobayashi scheme with the source message being partially decoded at each transmitter and obtained the achievable rate within a constant gap of the newly derived upper bound. Li *et al.* [15] investigated the diversity-multiplexing tradeoff of the delay-limited symmetric two-user fading IC and showed that the DF transmitter cooperations can notably improve the diversity performance. However, the above works assumed ideal self-interference (SI) cancellation at these FD transmitters and thus neglected the effect of the residual SI.

In this paper, we consider a Gaussian IC with two transmitters sending two messages to two receivers, respectively[1]. As explored in our previous study [19], transmitter cooperations can be constructed by in-band FD communications: The transmitters are capable of transmitting and receiving at the same frequency band; the transmitter cooperations and transmitter-receiver communications are with the same frequency band; and the SI cancellation at the transmitters is imperfect and the residual SI is thus remained in the received signal. Based on the above setup, we adopt the amplify-and-forward (AF) scheme at each transmitter to forward the received signals from its counterpart, which provides a much simpler analog receiver processing strategy compared with the conventional DF scheme. Moreover, to improve the achievable rates of the considered system, a repeated transmission strategy is adopted and each transmitter sends its own message twice in two successive time slots: At the first time slot, the two transmitters "exchange" their transmit signals, whereas the one from its counterpart may be corrupted by the residual SI (introduced by the imperfect SI cancellation) and the additive noise; at the second time slot, each transmitter sends its own message again as well as the received signal from its counterpart in the first time slot to form certain "common message" and transmitter beamforming is then applied to achieve higher power gain [20]. Meanwhile, since the residual SI and the additive noise at the transmitters cannot be eliminated, it will accumulate over time and thereby possibly degrades the system performance. The main contributions of this paper are summarized as follows:

1) First, an in-band FD transmitter cooperations framework with the AF scheme is proposed for the Gaussian IC and the equivalent channel model is analyzed. In particular, each transmitter first sends its own codewords and receives a corrupted version from its counterpart; then, after the SI cancellation and other processes, these two signals are obtained by both the transmitters and sent again to the destinations with properly designed transmission parameters that control the transmitter beamforming. At each receiver, we show that the channel for the considered system is equal to a two-tap MAC channel, with the additive noise determined by the accumulated residual self-interference and noise (ARIN).
2) Then, we investigate the statistics of the ARIN. As time passes, the ARIN at each transmitter is shown to form a Markov process, and when the time slot tends to infinity, the ARIN approaches a stationary state when certain convergence conditions for the transmission parameters are satisfied. When the stationary state is achieved, the covariance of the ARIN is also computed.
3) Finally, under the stationary state, we derive the average achievable rate regions for the IC based on the joint decoding and single-user decoding schemes. To characterize these regions, we adopt a two-stage iterative algorithm: First, the covariance matrix for the interference and noise of the equivalent two-tap channel are fixed, as well as the power of the ARIN, and then the transmission parameters are optimized by utilizing a two-step iterative method, with each step alternatively optimizing each transmitter's transmission parameters using a semidefinite relaxation (SDR) approach; next, the covariance matrix and the power of the ARIN are updated with the obtained transmission parameters.

[1] Since the signal processing and optimization methods for the single-user decoding scheme are similar to the ones for the joint decoding scheme, we omit the analysis of the single-user decoding scheme for simplicity.

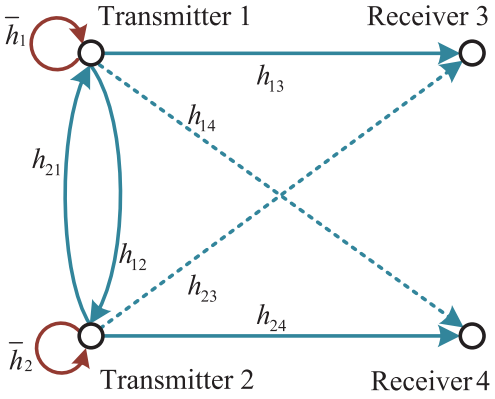


FIGURE 1. Interference channel with two full-duplex transmitters.

The remain of the paper is organized as follows: Section II introduces the system model; Section III analyzes the statistics of the ARIN, studies the rate regions for the two joint decoding schemes, and then proposes an efficient algorithm to compute the transmission parameters; Section IV presents some simulation results to validate our analysis; and Section V presents the conclusions.

Notations: $\log(x)$ and $|x|$ denote the base-2 logarithm and 2-norm of x. C^* and $\Re(C)$ denote the conjugate and real part of a complex number C. $\mathbf{M}^T$, $\mathbf{M}^H$, $\mathrm{Tr}(\mathbf{M})$, and $\mathrm{Rank}(\mathbf{M})$ denote the transpose, conjugate transpose, trace, and rank of matrix $\mathbf{M}$. $\mathrm{Diag}(x_1, x_2, \cdots, x_n)$ denotes a diagonal matrix with $x_1, \cdots, x_n$ being the diagonal elements. $\mathbf{M} \succeq 0$ means that $\mathbf{M}$ is a positive semidefinite matrix. In this paper, we denote j as index number of the transmitter $j, j = 1, 2$.

II. SYSTEM MODEL

In this paper, a Gaussian IC is considered as shown in Fig. 1, where two transmitters can cooperatively transmit messages and the two receivers decode the messages from both the transmitters. Transmission cooperations are established via in-band FD communications between the two transmitters: The transmitters are capable to simultaneously transmit and receive at the same frequency band, and each transmitter can receive the signal from its counterpart as well as that from itself, which is treated as SI and needs to be cancelled before further processions. The links between the two transmitters share the same frequency band as the other links between the transmitters and the receivers. Moreover, the channel coefficients among the transmitters and receivers are fixed after channel realizations and exactly known to the transmitters and receivers. In this paper, we consider the transmissions of N source messages at each transmitter, which are encoded separately into codewords $x_j(1), x_j(2), \cdots, x_j(N)$. The transmissions of N pairs of codewords are over $N+1$ time slots and the asymptotic performance of the considered system is considered as $N \to \infty$.

With the above setup, the transmissions and receptions for the considered IC are described in Fig. 2, where transmitter $\bar{j}$ denotes the counterpart of transmitter $j, j = 1, 2$. At the first time slot, transmitter j sends signal $t_j(1) = x_j(1)$ to its counterpart and the two receivers. Meanwhile, transmitter j receives the signals $t_j(1)$ and $t_{\bar{j}}(1)$ from itself and transmitter $\bar{j}$, respectively. For the received signal $r_j(1) = t_j(1) + t_{\bar{j}}(1) + n_j(1)$, $t_j(1)$ with much larger power compared to that of $t_{\bar{j}}(1)$ is treated as SI and needs to be cancelled, while part of it will still be remained due to the imperfection of the SI cancellation. After the cancellation of $t_j(1)$, the AF scheme is adopted to forward the residual signal $y_j(1) = x_{\bar{j}}(1) + \hat{y}_j(1)$, where $\hat{y}_j(1)$ is the residual SI and noise, to both receivers at the second time slot. Obviously, $\hat{y}_j(1)$ will be retained in the transmitted signal at the rest of the N time slot, and we call this phenomena as ARIN accumulation. At the second time slot, transmitter j sends its own new signal $x_j(2)$ and its previous $x_j(1)$ again, as well as $y_j(1)$. It is easy to observe that at the second time slot, both transmitters send $x_1(1)$ and $x_2(1)$, and potential power gain [20] can be achieved by properly designing the transmission parameters for transmitting these signals. In this paper, we assume the channel state information is exactly known after some channel estimation processes. In the following subsections, the signal transmissions and receptions for the i-th time slot are introduced, and the corresponding signal model is derived.

A. TRANSMISSION AT THE TRANSMITTERS

At the i-th time slot, transmitter j sends signal $t_j(i)$ to transmitter $\bar{j}$ and the two receivers, which consists of three parts: one new signal $x_j(i)$ sent for the first time at the current time slot, one old signal $x_j(i-1)$ also sent at the previous time slot, and the residual signal $y_j(i-1)$ after the SI cancellation of transmitter j at the $(i-1)$-th time slot. Thus, the transmitted signal $t_j(i)$ at transmitter j is given as

$$t_j(i) = w_{j1}x_j(i) + w_{j2}x_j(i-1) + w_{j3}y_j(i-1), \quad j = 1, 2, \quad (1)$$

where w_{j1}, w_{j2}, and w_{j3} are the complex transmission parameters to be designed in this paper. Signals $x_j(i)$ and $x_j(i-1)$ are with unit power, i.e., $\mathbb{E}[|x_j(i)|^2] = \mathbb{E}[|x_j(i-1)|^2] = 1$. The power of signal $y_j(i-1)$ is denoted as p_j^t, i.e., $\mathbb{E}[|y_j(i-1)|^2] = p_j^t$. Thus, the transmitted signal $t_j(i)$ satisfies a power constraint with the power budget P_j at transmitter j, i.e.,

$$\mathbb{E}\left(|t_j(i)|^2\right) = |w_{j1}|^2 + |w_{j2}|^2 + |w_{j3}|^2 p_j^t \le P_j. \quad (2)$$

B. RECEPTION AT THE TRANSMITTERS

At the end of the i-th time slot, transmitter j receives signal $t_j(i)$ from itself and signal $t_{\bar{j}}(i)$ from transmitter $\bar{j}$. Thus, the received signal $r_j(i)$ at transmitter j is given as

$$r_j(i) = h_{\bar{j}j}t_{\bar{j}}(i) + \bar{h}_j t_j(i) + n_j(i), \quad (3)$$

where $h_{\bar{j}j}$ denotes the channel coefficient from transmitter j to transmitter $\bar{j}$, $\bar{h}_j$ denotes the SI channel coefficient for transmitter j, and $n_j(i)$ is the independent and identically distributed (i.i.d.) complex circularly symmetric Gaussian (CSCG) noise with zero mean and variance σ_n^2.

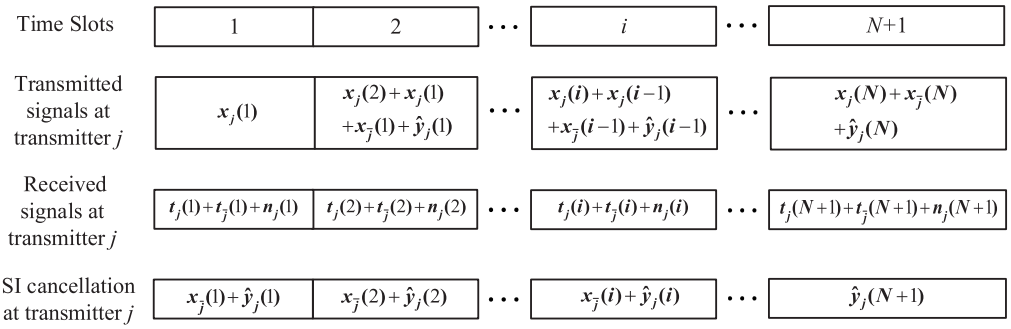


FIGURE 2. Transmissions and receptions of Gaussian IC over $N+1$ time slots.

By substituting (1) into (3), $r_j(i)$ can be rewritten as

$$r_j(i) = h_{\bar{j}j} w_{\bar{j}1} x_{\bar{j}}(i) + s_{j1}(i) + s_{j2}(i) + s_{j3}(i) + s_{jN}(i) + n_j(i), \quad (4)$$

with

$$s_{j1}(i) = \bar{h}_j t_j(i), \quad (5)$$

$$s_{j2}(i) = h_{\bar{j}j} h_{j\bar{j}} w_{\bar{j}3} w_{j1} x_j(i-1), \quad (6)$$

$$s_{j3}(i) = h_{\bar{j}j} w_{\bar{j}2} x_{\bar{j}}(i-1), \quad (7)$$

$$s_{jN}(i) = h_{\bar{j}j} w_{\bar{j}3} \hat{y}_{\bar{j}}(i-1), \quad (8)$$

where $x_{\bar{j}}(i)$ is from transmitter $\bar{j}$, $s_{j1}(i)$ denotes the SI directly from transmitter j, $s_{j2}(i)$ is a distorted version of signal $x_j(i-1)$ (initially transmitted from transmitter j to transmitter $\bar{j}$ at the $(i-1)$-th time slot and then forwarded by transmitter $\bar{j}$ at the i-th time slot), $s_{j3}(i)$ is signal $x_{\bar{j}}(i-1)$ sent from transmitter $\bar{j}$ at the i-th time slot, and $s_{jN}(i)$ is the ARIN from the previous time slots.

C. SI CANCELLATION AT THE TRANSMITTERS

From (4), we observe that: Signal $s_{j1}(i)$ is the SI from transmitter j itself and can be partially cancelled; signal $s_{j2}(i)$ is actually the signal $x_j(i-1)$, which is known to transmitter j and thus can be cancelled; and signal $s_{j3}(i)$ can be estimated from signal $y_j(i-1)$ after SI cancellation at the previous time slot and thus can also be partially cancelled (this issue will be discussed later in this subsection). Since the residual interferences are dominated by the ARIN with overwhelming power, the residual power of signal $s_{j2}(i)$ is rather smaller and can be neglected after SI cancellation. Hence, we denote the signal after the SI cancellation as $y_j(i)$, which is composed of the desired signal $x_{\bar{j}}(i)$ and the ARIN, i.e.,

$$y_j(i) = h_{\bar{j}j} w_{\bar{j}1} x_{\bar{j}}(i) + \hat{y}_j(i), \quad (9)$$

where $\hat{y}_j(i)$ is the sum of the residual parts of $s_{j1}(i)$ and $s_{j3}(i)$ denoted as $\hat{s}_{j1}(i)$ and $\hat{s}_{j3}(i)$, respectively, the ARIN $s_{jN}(i)$ from the previous time slot, and the CSCG noise $n_j(i)$, i.e.,

$$\hat{y}_j(i) = \hat{s}_{j1}(i) + \hat{s}_{j3}(i) + s_{jN}(i) + n_j(i). \quad (10)$$

As commonly implemented by SI cancellation process [21], the residual SI $\hat{s}_{j1}(i)$ is modeled as CSCG noise with zero mean and variance $\hat{P}_j$, where $\hat{P}_j$ is treated as a specific residual power, and is independent across different transmitters and different time slots.

Now, we introduce how to partially cancel signal $s_{j3}(i)$ and derive the expression for its residual signal $\hat{s}_{j3}(i)$. From (9), it is easy to see that $x_{\bar{j}}(i-1)$ can be estimated by $\frac{1}{h_{\bar{j}j} w_{\bar{j}1}} y_j(i-1)$, and then signal $s_{j3}(i)$ can be estimated by the processed signal $y_j(i-1)$ at the $(i-1)$-th time slot. Thus, $\hat{s}_{j3}(i)$ is given by

$$\begin{aligned} \hat{s}_{j3}(i) &= s_{j3}(i) - \frac{w_{\bar{j}2}}{w_{\bar{j}1}} y_j(i-1) \\ &= h_{\bar{j}j} w_{\bar{j}2} x_{\bar{j}}(i-1) - \frac{w_{\bar{j}2}}{w_{\bar{j}1}} y_j(i-1) \\ &= -\frac{w_{\bar{j}2}}{w_{\bar{j}1}} \hat{y}_j(i-1). \end{aligned} \quad (11)$$

From (11), we observe that signal $\hat{s}_{j3}(i)$ is determined by the ARIN $\hat{y}_j(i-1)$ from the previous time slot, which is independent with signal $\hat{s}_{j1}(i)$ and noise $n_j(i)$. By substituting (8) and (11) into (10), it follows

$$\hat{y}_j(i) = \alpha_{\bar{j}} \hat{y}_j(i-1) + h_{\bar{j}j} w_{\bar{j}3} \hat{y}_{\bar{j}}(i-1) + f_j(i), \quad (12)$$

where $f_j(i) = \hat{s}_{j1}(i) + n_j(i)$ and $\alpha_{\bar{j}} = -\frac{w_{\bar{j}2}}{w_{\bar{j}1}}$. In particular, when $w_{\bar{j}1} = 0$, $\alpha_{\bar{j}}$ is defined as 0[2].

Remark 1: From (12), we observe that:

- The ARIN $\left[\hat{y}_1(i),\ \hat{y}_2(i)\right]$ at the i-th time slot is related to the ARIN $\left[\hat{y}_1(i-1),\ \hat{y}_2(i-1)\right]$ at the $(i-1)$-th time slot and also affected by random variables $f_1(i)$ and $f_2(i)$. Therefore, the ARIN process over time forms a Markov process [22].
- From (1) and (9), it can be observed that at the i-th time slot, each transmitter sends both transmitters' signals $x_1(i-1)$ and $x_2(i-1)$. Thus, possible power gain can be achieved by properly designing the transmission parameters w_{j1}, w_{j2}, and w_{j3}. We should point out that since there are two receivers and the channel coefficients for the two MACs are not the same, we cannot simultaneously achieve the maximum power gains at both the receivers.

[2] When $w_{\bar{j}1} = 0$, transmitter $\bar{j}$ does not transmit its own messages, and thus $w_{\bar{j}2}$ is equal to 0. Hence, signal $s_{j3}(i)$ consisting of transmitter $\bar{j}$'s signal $x_{\bar{j}}(i-1)$ is also 0, as well as the residual part $\hat{s}_{j3}(i)$.

D. RECEPTION AT THE RECEIVERS

Receiver k ($k = 3, 4$, and receiver $\bar{k}$ denotes the counterpart of receiver$k$.) receives the signals from both the transmitters, and the received signal $r_k(i)$ is given as

$$r_k(i) = \sum_{j=1}^{2} h_{jk} t_j(i) + n_k(i), \quad k = 3, 4, \tag{13}$$

where h_{jk} is the channel coefficient from transmitter j to receiver k, and $n_k(i)$ is the CSCG noise with zero mean and variance σ_n^2. By substituting (1) and (9) into (13), it follows

$$r_k(i) = \sum_{j=1}^{2} \begin{pmatrix} \left(h_{jk} w_{j2} + h_{\bar{j}k} h_{j\bar{j}} w_{\bar{j}3} w_{j1}\right) x_j(i-1) \\ + h_{jk} w_{j1} x_j(i) + v_k(i) \end{pmatrix}, \tag{14}$$

with $v_k(i) = \sum_{j=1}^{2} h_{jk} w_{j3} \hat{y}_j(i-1) + n_k(i)$.

Remark 2: From (14), it is observed that for each receiver, the equivalent channel model is very similar to the two-tap MAC [23]. Due to the one-block delay generated by the forwarding operation at the transmitters, signal $x_j(i-1)$ at the $(i-1)$-th time slot will interfere signal $x_j(i)$ at the i-th time slot, which may cause some trouble in decoding at the receivers. On the other hand, the additive noise $v_k(i)$ contains ARIN from both the transmitters, where the two ARIN $\hat{y}_1(i-1)$ and $\hat{y}_2(i-1)$ are Markov processes according to (12). Hence, the noises $\{v_k(i)\}$ across different time slots are correlated and their correlation will make the decoding at each receiver more complicated.

Remark 3: From (14), it can be observed that the transmission of the N pairs of block messages will cost a total of $N + 1$ time slots. In this paper, we consider the case that N tends to infinity; thus, the one-block delay generated by the AF operation at the transmitters can be neglected. In the next section, we will further show that the statistics of the ARIN at the i-th time slot will converge to a stationary state, as well as the achievable rate pair for the two transmitted messages, as both N and i go to infinity. Thus, the average achievable rate pair of the considered IC with transmitter cooperations is given by that under this stationary state.

III. ACHIEVABLE RATE REGIONS

In this section, we focus on the joint decoding scheme, where each receiver decodes both the source messages, and characterize the corresponding achievable rate regions for the considered IC with transmitter cooperations. First, we analyze the statistics of the ARIN.

A. STATISTICS OF ARIN

In this subsection, the statistics of the ARIN at the stationary state, i.e., $i \to \infty$, is first computed. Define the ARIN vector at the i-th time slot as $\hat{\mathbf{Y}}(i) = \left[\hat{y}_1(i),\ \hat{y}_2(i)\right]^T$, and (12) can be rewritten as a matrix form

$$\hat{\mathbf{Y}}(i) = \mathbf{A}\hat{\mathbf{Y}}(i-1) + \mathbf{F}(i), \tag{15}$$

where $\mathbf{A}$ is a two by two matrix

$$\mathbf{A} = \begin{bmatrix} \alpha_2 & h_{21} w_{23} \\ h_{12} w_{13} & \alpha_1 \end{bmatrix}, \tag{16}$$

and $\mathbf{F}(i) = [f_1(i),\ f_2(i)]^T$, with $f_j(i) = \hat{s}_{j1}(i) + n_j(i)$ being defined in (12). Then, by mathematical induction, we can easily obtain the following proposition, whose proof is omitted due to space limitation.

Proposition 1: The ARIN vector $\hat{\mathbf{Y}}(i)$ can be recursively computed as

$$\hat{\mathbf{Y}}(i) = \sum_{n=0}^{i-1} \mathbf{A}^n \mathbf{F}(i-n). \tag{17}$$

From the above proposition, matrix $\mathbf{A}$ determines the speed of the accumulation of the ARIN: The larger the eigenvalue of $\mathbf{A}$ is, the faster the power of the ARIN increases. Now, with the above analysis, we arrive at the main result about the statistics of the ARIN.

Proposition 2: When N and $i \to \infty$, the covariance of the ARIN $\hat{\mathbf{Y}}(i)$ defined in (17) is computed as the following two cases:

1) Case I: Matrix $\mathbf{A}$ has two different eigenvalues λ_1 and λ_2, $\lambda_1 \neq \lambda_2$. Thus, according to Theorem 7.1.6 in [24], matrix $\mathbf{A}$ can be decomposed as

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{P}^{-1}, \tag{18}$$

where $\mathbf{P} = \begin{bmatrix} p_{11}\ p_{12} \\ p_{21}\ p_{22} \end{bmatrix}$ is a two by two invertible matrix. Then, when $|\lambda_j| < 1$ and $i, k \to \infty$, the covariance matrix of $\hat{\mathbf{Y}}(i)$ is asymptotically given as (19), as shown at the top of the next page, where $(x)^+ = \max(x, 0)$, $\varpi = i - k$, and

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{12}^* & f_{22} \end{bmatrix} = \mathbf{P}^{-1} \begin{bmatrix} \hat{P}_1 + \sigma_n^2 & 0 \\ 0 & \hat{P}_2 + \sigma_n^2 \end{bmatrix} \left(\mathbf{P}^{-1}\right)^H, \tag{20}$$

where $\hat{P}_j$, $j = 1, 2$, is the power of the residual SI $\hat{s}_{j1}(i)$, and σ_n^2 is the power of the CSCG noise $n_j(i)$.

2) Case II: Matrix $\mathbf{A}$ has two identical eigenvalues, i.e., $\lambda_1 = \lambda_2 = \lambda$. Thus, according to Theorem 7.1.9 in [24], matrix $\mathbf{A}$ can be decomposed as a *Jordan canonical form* with an invertible matrix $\mathbf{P}$, i.e.,

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \mathbf{P}^{-1}. \tag{21}$$

Similarly, when $|\lambda| < 1$ and $i, k \to \infty$, the covariance matrix of $\hat{\mathbf{Y}}(i)$ is asymptotically given as (22), as shown at the top of the next page, where

$$c_{11} = \frac{f_{11}}{1 - |\lambda|^2} + \frac{f_{12}\lambda}{(1 - |\lambda|^2)^2} + \frac{f_{12}^* \lambda^*}{(1 - |\lambda|^2)^2} + \frac{f_{22}(1 + |\lambda|^2)}{(1 - |\lambda|^2)^3}, \tag{23}$$

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbf{P}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}^{(\varpi)^+}\begin{bmatrix}\frac{1}{1-|\lambda_1|^2}f_{11} & \frac{1}{1-\lambda_1\lambda_2^*}f_{12}\\ \frac{1}{1-\lambda_1^*\lambda_2}f_{12}^* & \frac{1}{1-|\lambda_2|^2}f_{22}\end{bmatrix}\begin{bmatrix}\lambda_1^* & 0\\ 0 & \lambda_2^*\end{bmatrix}^{(-\varpi)^+}\mathbf{P}^H, \tag{19}$$

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbf{P}\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}^{(\varpi)^+}\begin{bmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{bmatrix}\begin{bmatrix}\lambda^* & 0\\ 1 & \lambda^*\end{bmatrix}^{(-\varpi)^+}\mathbf{P}^H, \tag{22}$$

$$\mathbb{E}\left[|\hat{y}_1(i)|^2\right] = \frac{1}{1-|\lambda_1|^2}|p_{11}|^2 f_{11} + \Re\left(p_{11}^* p_{12}\frac{1}{1-\lambda_1^*\lambda_2}f_{12}^*\right) + \frac{1}{1-|\lambda_2|^2}|p_{12}|^2 f_{22}, \tag{26}$$

$$\mathbb{E}\left[|\hat{y}_2(i)|^2\right] = \frac{1}{1-|\lambda_1|^2}|p_{21}|^2 f_{11} + \Re\left(p_{21}^* p_{22}\frac{1}{1-\lambda_1^*\lambda_2}f_{12}^*\right) + \frac{1}{1-|\lambda_2|^2}|p_{22}|^2 f_{22}. \tag{27}$$

$$c_{12} = \frac{f_{12}}{1-|\lambda|^2} + \frac{f_{22}\lambda^*}{(1-|\lambda|^2)^2}, \tag{24}$$

$$c_{21} = c_{12}^*, \quad c_{22} = \frac{f_{22}}{1-|\lambda|^2}. \tag{25}$$

Proof: Please see Appendix A. ∎

Remark 4: As mentioned in the previous section, the power of the ARIN will accumulate as time goes. Notice that if we do not consider the power constraint in (2), when $|\lambda_j| \geq 1$ and i tends to infinity, the power of the ARIN $\hat{y}_j(i)$ will also tend to infinity and submerge the desired signals. On the other hand, when $|\lambda_j| < 1$, the power of the ARIN will converge to a finite constant. Hence, we only consider the stationary state as $i \to \infty$ with $\max(|\lambda_1|, |\lambda_2|) < 1$.

Corollary 1: According to Proposition 2, when $\max(|\lambda_1|, |\lambda_2|) < 1$ and $i \to \infty$, the power of the ARIN is given as

- Case I: (26) and (27), as shown at the top of this page.
- Case II:

$$\mathbb{E}\left[|\hat{y}_1(i)|^2\right] = |p_{11}|^2 c_{11} + \Re\left(p_{11}^* p_{12} c_{12}^*\right) + c_{22}|p_{12}|^2, \tag{28}$$

$$\mathbb{E}\left[|\hat{y}_2(i)|^2\right] = |p_{21}|^2 c_{11} + \Re\left(p_{22} p_{21}^* c_{12}^*\right) + c_{22}|p_{22}|^2. \tag{29}$$

Based on the above results, we can explicitly compute the sum power constraint in (2). From (9) and (10), the power of signal $y_j(i)$ given in (9) can be computed as

$$\begin{aligned} p_j^t &= \mathbb{E}\left[|y_j(i)|^2\right] = \mathbb{E}\left(|h_{\bar{j}\bar{j}} w_{\bar{j}1} x_{\bar{j}}(i)|^2\right) + \mathbb{E}\left(|\hat{y}_j(i)|^2\right) \\ &= g_{\bar{j}\bar{j}}|w_{\bar{j}1}|^2 + \hat{a}_j, \end{aligned} \tag{30}$$

where $\hat{a}_j = \mathbb{E}\left(|\hat{y}_j(i)|^2\right)$ is given in Corollary 1 and $g_{\bar{j}\bar{j}} = |h_{\bar{j}\bar{j}}|^2$. Hence, by (30), the sum power constraint in (2) can be rewritten as

$$|w_{j1}|^2 + |w_{j2}|^2 + |w_{j3}|^2\left(g_{\bar{j}\bar{j}}|w_{\bar{j}1}|^2 + \hat{a}_j\right) \leq P_j. \tag{31}$$

Therefore, we define the transmitter power region as

$$\mathcal{P} = \left\{(w_{j1}, w_{j2}, w_{j3}) : (31),\ j = 1, 2\right\}. \tag{32}$$

B. JOINT DECODING SCHEMES

From the two-tap channel model in (14), the optimal decoder for the considered IC with transmitter cooperations is simultaneously decoding all the messages transmitted at all the time slots. However, this optimal decoding scheme requires the receivers to buffer all the received signals and the decoding complexity becomes exponentially high as N increases, which makes this decoder too complicated to be realized in practice. As a tradeoff for the decoding complexity, we mainly focus on some suboptimal while simple decoding schemes, which require each receiver to sequentially decode only the messages of the codewords $x_1(i)$ and $x_2(i)$ transmitted at the i-th time slot. From (14), it is observed that both $r_k(i)$ and $r_k(i+1)$ contain the codewords $x_1(i)$ and $x_2(i)$. As such, in order to achieve higher received signal-to-noise ratio (SNR) for the decoding of these two messages, we need to utilize both $r_k(i)$ and $r_k(i+1)$ to decode them. However, the codewords $x_1(i)$ and $x_2(i)$ are interfered by signal $x_j(i-1)$ or $x_j(i+1)$ (see (14)), depending on the decoding order for the transmitted signals. Therefore, we consider the following two decoding schemes [25]:

1) Forward decoding scheme: Each receiver decodes the source messages from the first time slot to the last time slot, and thus, the receiver only needs to buffer the signals at two successive time slots. At the $i+1$-th time slot, it is assumed that $x_j(1), \cdots, x_j(i-1)$ have been successfully decoded, and thus $x_1(i-1)$ and $x_2(i-1)$ can be completely eliminated from $r_3(i)$ and $r_4(i)$. Then, each receiver tries to decode $x_1(i)$ and $x_2(i)$ transmitted at the i-th time slot, and treats $x_1(i+1)$ and $x_2(i+1)$ as interference. Based on (14), the equivalent channel input and output relationship for the forward decoding is given as

$$\mathbf{R}_k = \mathbf{H}_{jk} x_j(i) + \mathbf{H}_{\bar{j}k} x_{\bar{j}}(i) + \mathbf{Z}_k^f, \tag{33}$$

where $\mathbf{R}_k = [r_k(i+1), r_k(i)]^T$, $\mathbf{H}_{jk} = \left[b_{jk} + c_{\bar{j}k},\ a_{jk}\right]^T$, $\mathbf{Z}_k^f = \left[z_{k1}^f,\ z_{k2}^f\right]^T$, and

$$z_{k1}^f = n_k(i+1) + \sum_{j=1}^{2} a_{jk} x_j(i+1) + e_{jk}\hat{y}_j(i), \tag{34}$$

$$z_{k2}^f = n_k(i) + \sum_{j=1}^{2} e_{jk}\hat{y}_j(i-1), \tag{35}$$

with $a_{jk}=h_{jk}w_{j1}$, $b_{jk}=h_{jk}w_{j2}$, $e_{jk}=h_{jk}w_{j3}$, and $c_{jk}=h_{jk}h_{\bar{j}j}w_{j3}w_{\bar{j}1}$.

2) Backward decoding scheme: Each receiver decodes the source messages from the last time slot to the first time slot, and thus the receiver needs to buffer all the received signals at all the time slots. When each receiver tries to decode $x_1(i)$ and $x_2(i)$, it is assumed that $x_j(i+1), \cdots, x_j(N)$ have already been successfully decoded, and thus $x_1(i+1)$ and $x_2(i+1)$ can be completely eliminated from $r_3(i+1)$ and $r_4(i+1)$. Then, $x_1(i)$ and $x_2(i)$ can be decoded at each receiver, and $x_1(i-1)$ and $x_2(i-1)$ are treated as interference. Based on (14), the equivalent channel input and output relationship for the backward decoding is given as

$$\mathbf{R}_k = \mathbf{H}_{jk}x_j(i) + \mathbf{H}_{\bar{j}k}x_{\bar{j}}(i) + \mathbf{Z}_k^b, \tag{36}$$

where $\mathbf{Z}_k^b = \left[z_{k1}^b,\ z_{k2}^b\right]^T$, with

$$z_{k1}^b = n_k(i+1) + \sum_{j=1}^{2} e_{jk}\hat{y}_j(i), \tag{37}$$

$$z_{k2}^b = n_k(i) + \sum_{j=1}^{2} (b_{jk} + c_{\bar{j}k})x_j(i-1) + e_{jk}\hat{y}_j(i-1). \tag{38}$$

Remark 5: From (33) and (36), we observe that the equivalent channel models for the two decoding schemes are quite similar: 1) The channel gains for the desired information $x_1(i)$ and $x_2(i)$ are the same; 2) the equivalent channel coefficient $\mathbf{H}_{j,k}$ comes from the transmission parameters w_{j1}, w_{j2}, and w_{j3}, $j=1,2$, and properly designed transmission parameters can achieve different requirements for the transmission rates of the two transmitters; and 3) the additive noises for the two schemes are different, and significantly affect the performance of the two schemes.

Now, we are ready to compute the covariance matrix $\mathbf{Q}_k$ of the additive noise $\mathbf{Z}_k$. Here, without introducing any confusion, we ignore the superscript f and b for the variables $\mathbf{Q}_k$ and $\mathbf{Z}_k$, and will add them to specify the forward and backward decoding schemes when necessary.

1) For the forward decoding scheme, it follows

$$\begin{aligned}\mathbf{Q}_k &= \mathbf{Q}_k^f = \mathbb{E}\left[\mathbf{Z}_k^f \cdot \left(\mathbf{Z}_k^f\right)^H\right] \\ &= \begin{bmatrix} \sigma_n^2+|a_{1k}|^2+|a_{2k}|^2+Y_{i,i} & Y_{i,i-1} \\ Y_{i,i-1}^* & \sigma_n^2+Y_{i-1,i-1}\end{bmatrix},\end{aligned} \tag{39}$$

where $Y_{i,j}$ is calculated by the statistics of the ARIN obtained in (19) and (22),

$$Y_{i,j} = \begin{bmatrix} e_{1k} & e_{2k}\end{bmatrix}\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(j)^H\right)\begin{bmatrix} e_{1k} & e_{2k}\end{bmatrix}^H. \tag{40}$$

2) For the backward decoding scheme, it follows (41) and (42), as shown at the top of the next page.

In contrast to joint decoding scheme, single-user decoding scheme allows each receiver to decode the desired transmitter's messages while treating the other transmitter's signals simply as noises. Here, due to space limitation, we omit the analysis for the single-user decoding scheme in which the aforementioned forward and backward decoding schemes can be easily applied to this scheme with the covariance matrix $\mathbf{Q}_k$ being calculated in similar ways (39), (41).

C. ACHIEVABLE RATE REGIONS

Based on the equivalent parallel channel models given in (33) and (36) for the joint forward and backward decoding schemes, the achievable rate regions $\mathcal{C}(\mathcal{P})$ is asymptotically obtained by the results for the conventional MAC [20], [26], i.e., (43), as shown at the top of the next page.

For the single-user decoding scheme, the rate region [13], [27] is characterized as (44), as shown at the top of the next page.

In this paper, we adopt the rate profile approach [28] to characterize the rate regions defined in (43). Choose a rate profile parameter α, $0 \le \alpha \le 1$, and let $\bar{\alpha} = 1-\alpha$. Then, we define the sum rate as $R_{sum} = R_1 + R_2$, and set $R_1 = \alpha R_{sum}$ and $R_2 = \bar{\alpha}R_{sum}$. For a fixed α, we solve the following sum rate maximization problem intended for the joint decoding scheme (similarly for the single-user decoding scheme by replacing the rate constraints with (44)) subject to the rate constraints in (43), the power constraint in (32), and the forwarding constraint $\max(|\lambda_1|, |\lambda_2|) < 1$, i.e.,

$$\begin{aligned}\max_{R_{sum},\{w_{jl}\}}\ & R_{sum} \\ \text{s. t. }\ & \alpha R_{sum} \le \log\left(1+\mathbf{H}_{1k}^H\mathbf{Q}_k^{-1}\mathbf{H}_{1k}\right), \\ & \bar{\alpha}R_{sum} \le \log\left(1+\mathbf{H}_{2k}^H\mathbf{Q}_k^{-1}\mathbf{H}_{2k}\right), \\ & R_{sum} \le \log\left(1+\mathbf{H}_{2k}^H\mathbf{Q}_k^{-1}\mathbf{H}_{2k}+\mathbf{H}_{1k}^H\mathbf{Q}_k^{-1}\mathbf{H}_{1k}\right), \\ & (w_{j1}, w_{j2}, w_{j3}) \in \mathcal{P}, \\ & \max(|\lambda_1|, |\lambda_2|) < 1, \quad k=3,4,\end{aligned} \tag{45}$$

where $R_{sum} \in \mathbb{R}$, $w_{jl} \in \mathbb{C}$, $\mathbb{R}$ is the set of all real values, and $\mathbb{C}$ is the set of all complex values. However, it is obvious that all the constraints in (45) are non-convex. Thus, in this paper, we turn to a suboptimal while efficient algorithm to solve problem (45).

D. ALGORITHM

We adopt an iterative method to approximately solve problem (45) which can be easily extended to single-user decoding scenario, and in each iteration, the following two steps are taken:

1) Fix $\mathbf{Q}_k$ and the power of the ARIN $\hat{a}_j$ defined in Corollary 1, and optimize the transmission parameters $\{w_{jl}\}$;
2) update $\mathbf{Q}_k$ and $\hat{a}_j$ with the obtained transmission parameters $\{w_{jl}\}$ in step 1).

For Step 2), the new $\mathbf{Q}_k$ can be computed using (39) and (41). Thus, in the sequel, we mainly focus on Step 1).

$$\mathbf{Q}_k = \mathbf{Q}_k^b = \mathbb{E}\left[\mathbf{Z}_k^b \cdot \left(\mathbf{Z}_k^b\right)^H\right] \tag{41}$$

$$= \begin{bmatrix} \sigma_n^2 + Y_{i,i} & Y_{i,i-1} \\ Y_{i,i-1}^* & \sigma_n^2 + |b_{1k} + c_{2k}|^2 + |b_{2k} + c_{1k}|^2 + Y_{i-1,i-1} \end{bmatrix}. \tag{42}$$

$$\mathcal{C}(\mathcal{P}) \triangleq \bigcup_{\substack{(w_{j1},w_{j2},w_{j3})\in\mathcal{P} \\ \max(|\lambda_1|,|\lambda_2|)<1}} \left\{ (R_1,R_2) \left| \begin{array}{l} R_1 \le \log\left(1 + \mathbf{H}_{1k}^H \mathbf{Q}_k^{-1} \mathbf{H}_{1k}\right) \\ R_2 \le \log\left(1 + \mathbf{H}_{2k}^H \mathbf{Q}_k^{-1} \mathbf{H}_{2k}\right) \\ R_1 + R_2 \le \log\left(1 + \mathbf{H}_{2k}^H \mathbf{Q}_k^{-1} \mathbf{H}_{2k} + \mathbf{H}_{1k}^H \mathbf{Q}_k^{-1} \mathbf{H}_{1k}\right) \end{array} \right., k = 3,4 \right\}. \tag{43}$$

$$\mathcal{C}(\mathcal{P}) \triangleq \bigcup_{\substack{(w_{j1},w_{j2},w_{j3})\in\mathcal{P} \\ \max(|\lambda_1|,|\lambda_2|)<1}} \left\{ (R_1,R_2) \left| \begin{array}{l} R_1 \le \log\left(1 + \mathbf{H}_{13}^H \mathbf{Q}_3^{-1} \mathbf{H}_{13}\right) \\ R_2 \le \log\left(1 + \mathbf{H}_{24}^H \mathbf{Q}_4^{-1} \mathbf{H}_{24}\right) \end{array} \right. \right\}. \tag{44}$$

Considering $\mathbf{H}_{jk}$ given in (33) and (36), it is easy to observe that the transmission parameters of the two transmitters are coupled together. Hence, we utilize a two-stage iterative method to optimize these transmission parameters: In each iteration, we fix one transmitter $\bar{j}$'s parameters $w_{\bar{j}1}$, $w_{\bar{j}2}$, and $w_{\bar{j}3}$, optimize transmitter j's parameters w_{j1}, w_{j2}, and w_{j3}, and repeat the aforementioned optimization process by exchanging the places of the two transmitters' parameters in the optimization problem. Now, without loss of generality, we show how to optimize w_{j1}, w_{j2}, and w_{j3} with the fixed $w_{\bar{j}1}$, $w_{\bar{j}2}$, and $w_{\bar{j}3}$, and the alternative optimization problem is similar. First, to simplify the notations, we define two variables $\mathbf{w}_{j1} \in \mathbb{C}^2$ and $\mathbf{w}_{j2} \in \mathbb{C}^2$ as follows

$$\mathbf{w}_{j1} = \begin{bmatrix} w_{j2} \\ w_{j1} \end{bmatrix}, \quad \mathbf{w}_{j2} = \begin{bmatrix} w_{j3} \\ 1 \end{bmatrix}, \tag{46}$$

where $\mathbb{C}^n$ is the set of all complex $n \times 1$ vectors. Then, matrix $\mathbf{H}_{jk}$ defined in (33) and (36) can be simplified as

$$\mathbf{H}_{jk} = \begin{bmatrix} h_{jk} w_{j2} + h_{\bar{j}k} h_{\bar{j}\bar{j}} w_{\bar{j}3} w_{j1} \\ h_{jk} w_{j1} \end{bmatrix} = \mathbf{G}_{\bar{j}k1} \mathbf{w}_{j1} = \mathbf{G}_{jk2} \mathbf{w}_{\bar{j}2}, \tag{47}$$

where

$$\mathbf{G}_{\bar{j}k1} = \begin{bmatrix} h_{jk} & h_{\bar{j}k} h_{\bar{j}\bar{j}} w_{\bar{j}3} \\ 0 & h_{jk} \end{bmatrix}, \tag{48}$$

$$\mathbf{G}_{jk2} = \begin{bmatrix} h_{\bar{j}k} h_{j\bar{j}} w_{j1} & h_{jk} w_{j2} \\ 0 & h_{jk} w_{j1} \end{bmatrix}. \tag{49}$$

Therefore, based on the above notations, we can formulate the transmission parameters design problem of transmitter j for problem (45) as

$$\begin{aligned} \max_{R_{sum},\{\mathbf{w}_{ij}\}} \quad & R_{sum} \\ \text{s. t.} \quad & 2^{\alpha R_{sum}} - 1 \le \mathbf{w}_{jj}^H \mathbf{T}_{\bar{j}kj} \mathbf{w}_{jj}, \\ & 2^{\bar{\alpha} R_{sum}} - 1 \le \mathbf{w}_{\bar{j}\bar{j}}^H \mathbf{T}_{\bar{j}kj} \mathbf{w}_{\bar{j}\bar{j}}, \\ & 2^{R_{sum}} - 1 \le \mathbf{w}_{\bar{j}\bar{j}}^H \mathbf{T}_{\bar{j}kj} \mathbf{w}_{\bar{j}\bar{j}} + \mathbf{w}_{jj}^H \mathbf{T}_{\bar{j}kj} \mathbf{w}_{jj}, \\ & \mathbf{w}_{j1}^H \mathbf{w}_{j1} + \left(\mathbf{w}_{j2}^H \mathbf{w}_{j2} - 1\right) m_j \le P_j, \\ & g_{\bar{j}\bar{j}} \mathbf{w}_{j1}^H \mathrm{Diag}(0,1) \mathbf{w}_{j1} \le \bar{P}_{\bar{j}}, \\ & \max(|\lambda_1|, |\lambda_2|) < 1, \quad k = 3,4, \end{aligned} \tag{50}$$

where $\bar{P}_{\bar{j}} = \frac{P_{\bar{j}} - \mathbf{w}_{j1}^H \mathbf{w}_{\bar{j}1}}{\mathbf{w}_{j2}^H \mathbf{w}_{\bar{j}2} - 1} - \hat{a}_{\bar{j}}$, $\mathbf{T}_{\bar{j}kj} = \mathbf{G}_{jkj}^H \mathbf{Q}_k^{-1} \mathbf{G}_{\bar{j}kj}$, $\mathbf{T}_{\bar{j}k\bar{j}} = \mathbf{G}_{\bar{j}k\bar{j}}^H \mathbf{Q}_k^{-1} \mathbf{G}_{\bar{j}k\bar{j}}$, $m_j = g_{j\bar{j}} |w_{\bar{j}1}|^2 + \hat{a}_j$ are all constants, with $g_{j\bar{j}} = |h_{j\bar{j}}|^2$ and $\hat{a}_j$ being given in (30). However, problem (50) is still non-convex: The rate constraints in (50) are non-convex since the matrices $\mathbf{T}_{\bar{j}kj}$ and $\mathbf{T}_{\bar{j}k\bar{j}}$ are indefinite Hermitian matrices; and the eigenvalue constraint $\max(|\lambda_1|, |\lambda_2|) < 1$ is also non-convex. In the following, we will show how to handle these constraints and efficiently solve problem (50).

1) EIGENVALUE CONSTRAINT APPROXIMATION

In the following Propositoin 3, we approxiamte the complicated eigenvalue constraint $\max(|\lambda_1|, |\lambda_2|) < 1$ with two simpler quadratical constraints, which might cause the loss of some feasible points of problem (50) and introduce some infeasible ones. In the sequel, we will further show how to deal with the drawbacks of this method.

Proposition 3: The constraint $\max(|\lambda_1|, |\lambda_2|) < 1$ in problem (50) can be approximated by

$$\mathbf{w}_{j1}^H \mathbf{E}_0 \mathbf{w}_{j1} \le 0, \tag{51}$$

$$g_{j\bar{j}} \mathbf{w}_{j2}^H \mathbf{E}_1 \mathbf{w}_{j2} \le 1, \tag{52}$$

where $\mathbf{E}_0 = \mathrm{Diag}(1, -1)$ and $\mathbf{E}_1 = \mathrm{Diag}(1, 0)$.

Here, we argue the plausibility of the proposed method. The above approximation in Proposition 3 can be adjusted by modifying the constant parameters on the right side of (51) and (52). For example, by considering two new constraints $\mathbf{w}_{j1}^H \mathrm{Diag}(1, -0.25) \mathbf{w}_{j1} \le 0$ and $g_{j\bar{j}} \mathbf{w}_{j2}^H \mathbf{E}_1 \mathbf{w}_{j2} \le 0.25$, which is equivalent to $|a_j| = |\frac{w_{j2}}{w_{j1}}| \le 0.5$ and $|h_{j\bar{j}} w_{j3}| \le 0.5$, we can show that the eigenvalues λ_1 and λ_2 satisfy

$$|h_{j\bar{j}} w_{j3}| \ge |\lambda_j + a_j| \tag{53}$$

$$\ge |\lambda_j| - |a_j|, \tag{54}$$

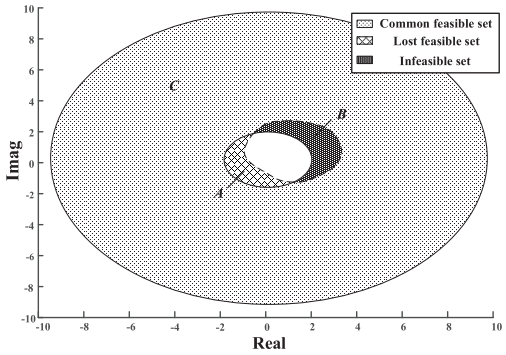


FIGURE 3. Feasible sets of w_{11} under the constraint $\max(|\lambda_1|, |\lambda_2|) < 1$ and Proposition 3.

where (53) is by Gershgorin circle theorem [24]. With (54), it follows $|\lambda_j| \leq |a_j| + |h_{j\bar{j}} w_{j3}| \leq 1$, and thus $\max(|\lambda_1|, |\lambda_2|) \leq 1$. Obviously, this approximation leads to a subset of the constraint $\max(|\lambda_1|, |\lambda_2|) \leq 1$, whereas it may cause the loss of too many feasible points of the original constraint. To reduce the effect of this issue, we relax the eigenvalue constraint as (51) and (52) (by simulations, the approximation given in (51) and (52) provides a good performance), while it may introduce some infeasible points. Now, we provide an example to describe this phenomenon. By choosing the parameters $w_{12} = -1.6549 - 1.1231i$, $w_{13} = -0.0023 + 0.0299i$, $w_{21} = -3.9919 - 3.01i$, $w_{22} = -0.9651 - 3.8818i$, $w_{23} = 0.0071 + 0.0039i$, $P_j = 20$ dB, $h_{12} = h_{21} = 10$, and $\hat{a}_j = 0$, we draw the sets of w_{11} under the constraint $\max(|\lambda_1|, |\lambda_2|) < 1$ and Proposition 3 in Fig. 3 (both of them are subject to the power constraint in (31)),
where area $A \cup C$ is the feasible set given by $\max(|\lambda_1|, |\lambda_2|) < 1$ and area $B \cup C$ is the feasible set given by Proposition 3. From Fig. 3, it is observed that the proposed approximation will lose the feasible set A and introduce infeasible set B, and leads the solution of problem (50) to be suboptimal and infeasible under certain cases (In the following part **(III)**, we will further show how to deal with these drawbacks.).

2) SEMIDEFINITE RELAXATION

Define variables $\mathbf{W}_{j1} = \mathbf{w}_{j1}\mathbf{w}_{j1}^H$ and $\mathbf{W}_{j2} = \mathbf{w}_{j2}\mathbf{w}_{j2}^H$, and the rank-1 constraint $\text{rank}(\mathbf{W}_{j1}) = \text{rank}(\mathbf{W}_{j2}) = 1$. Then, problem (50) after the approximation of Proposition 3 is equivalent to the following form

$$\begin{aligned}
\max_{R_{sum},\{\mathbf{W}_{ij}\}} \quad & R_{sum} \\
\text{s. t.} \quad & 2^{\alpha R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}kj}\mathbf{W}_{jj}), \\
& 2^{\bar{\alpha} R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}k\bar{j}}\mathbf{W}_{\bar{j}\bar{j}}), \\
& 2^{R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}kj}\mathbf{W}_{jj}) + \text{Tr}(\mathbf{T}_{\bar{j}k\bar{j}}\mathbf{W}_{\bar{j}\bar{j}}), \\
& \text{rank}(\mathbf{W}_{j1}) = \text{rank}(\mathbf{W}_{j2}) = 1, \ \text{Tr}\left(\mathbf{E}_2\mathbf{W}_{j2}\right) = 1, \\
& \mathbf{W}_{j1} \succeq 0, \ \mathbf{W}_{j2} \succeq 0, \ \left(\mathbf{W}_{j1}, \mathbf{W}_{j2}\right) \in \mathcal{P}_0, \\
& \text{Tr}(\mathbf{E}_0\mathbf{W}_{j1}) \leq 0, \ g_{\bar{j}\bar{j}}\text{Tr}(\mathbf{E}_1\mathbf{W}_{j2}) \leq 1, \quad k = 3, 4, \quad (55)
\end{aligned}$$

where $\mathcal{P}_0$ satisfies (56), as shown at the bottom of the next page, $\mathbf{E}_2 = \text{Diag}(0, 1)$, $\mathbf{W}_{ij} \in \mathbb{H}^2$, and $\mathbb{H}^n$ is the set of all complex $n \times n$ Hermitian matrices. It is noticed that since the second element of vector $\mathbf{w}_{j2}$ is 1, we set the $(2,2)^{th}$ element of matrix $\mathbf{W}_{j2}$ is 1, i.e., $\text{Tr}\left(\mathbf{E}_2\mathbf{W}_{j2}\right) = 1$.

It is observed that the rank-1 constraint constraint $\text{rank}(\mathbf{W}_{j1}) = \text{rank}(\mathbf{W}_{j2}) = 1$ in (55) is non-convex and all the other constraints are convex. Thus, we drop the non-convex rank-1 constraint and obtain a relaxed version of problem (55) as follows

$$\begin{aligned}
\max_{R_{sum},\{\mathbf{W}_{ij}\}} \quad & R_{sum} \\
\text{s. t.} \quad & 2^{\alpha R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}kj}\mathbf{W}_{jj}), \\
& 2^{\bar{\alpha} R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}k\bar{j}}\mathbf{W}_{\bar{j}\bar{j}}), \\
& 2^{R_{sum}} - 1 \leq \text{Tr}(\mathbf{T}_{\bar{j}kj}\mathbf{W}_{jj}) + \text{Tr}(\mathbf{T}_{\bar{j}k\bar{j}}\mathbf{W}_{\bar{j}\bar{j}}), \\
& \mathbf{W}_{j1} \succeq 0, \ \mathbf{W}_{j2} \succeq 0, \ \left(\mathbf{W}_{j1}, \mathbf{W}_{j2}\right) \in \mathcal{P}_0, \\
& \text{Tr}(\mathbf{E}_0\mathbf{W}_{j1}) \leq 0, \ g_{\bar{j}\bar{j}}\text{Tr}(\mathbf{E}_1\mathbf{W}_{j2}) \leq 1, \\
& \text{Tr}\left(\mathbf{E}_2\mathbf{W}_{j2}\right) = 1, \quad k = 3, 4 \quad (57)
\end{aligned}$$

Problem (57) is the SDR of problem (50) and can be effectively solved by the interior-point algorithm [29].

3) GAUSSIAN RANDOMIZATION

When the obtained solution of problem (57) is not rank 1 or not feasible for the eigenvalue constraint $\max(|\lambda_1|, |\lambda_2|) < 1$, we apply the following Gaussian randomization [30] to construct an approximated solution for problem (50). The main idea of this method is described as follows: First, obtain an optimal solution $\{\mathbf{W}_{ij}^*\}$ of problem (57); then, generate multiple random samples $\hat{\mathbf{w}}_{ij} \in \mathbb{C}^2$ according to the Gaussian distribution with zero mean and covariance matrix $\mathbf{W}_{ij}^*$, i.e., $\hat{\mathbf{w}}_{ij} \sim \mathcal{CN}(0, \mathbf{W}_{ij}^*)$, and set the second element of $\hat{\mathbf{w}}_{j2}$ as 1; next, adopt a scaling method to ensure all the generated samples satisfying the power constraint in (31): When the generated vectors exceed the maximum power budget in (31), we can shrink them as follows

$$\hat{\mathbf{w}}_{j1} = \frac{\hat{\mathbf{w}}_{j1}\sqrt{P_j}}{\sqrt{\hat{\mathbf{w}}_{j1}^H\hat{\mathbf{w}}_{j1} + \left(\hat{\mathbf{w}}_{j2}^H\hat{\mathbf{w}}_{j2} - 1\right)m_j}}, \quad (58)$$

$$\hat{\mathbf{w}}_{j1}(2) = \frac{\hat{\mathbf{w}}_{j1}(2)\sqrt{\bar{P}_{\bar{j}}}}{\sqrt{g_{\bar{j}\bar{j}}\hat{\mathbf{w}}_{j1}^H\mathbf{E}_2\hat{\mathbf{w}}_{j1}}}, \quad (59)$$

$$\hat{\mathbf{w}}_{j2} = \frac{\hat{\mathbf{w}}_{j2}\sqrt{P_j}}{\sqrt{\hat{\mathbf{w}}_{j1}^H\hat{\mathbf{w}}_{j1} + \left(\hat{\mathbf{w}}_{j2}^H\hat{\mathbf{w}}_{j2} - 1\right)m_j}}, \quad (60)$$

where $\hat{\mathbf{w}}_{j1}(2)$ is the second element of $\hat{\mathbf{w}}_{j1}$; next, drop the samples that are not feasible for the eigenvalue constraint $\max(|\lambda_1|, |\lambda_2|) < 1$; finally, among these feasible samples, choose one couple of vectors $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$, which maximize the sum rate $\hat{R}_{sum}$ defined as (61), as shown at the bottom of the next page, according to the rate constraints in (50).

TABLE 1. Algorithm 1: Gaussian randomization method for optimizing $\mathbf{w}_{j1}$ and $\mathbf{w}_{j2}$.

Input: the total sample number I, $\mathbf{Q}_k$, $\hat{a}_j$, $\mathbf{w}_{\bar{j}1}$, $\mathbf{w}_{\bar{j}2}$, and the power budget P_j

Output: $\hat{\mathbf{w}}_{j1}^*$, $\hat{\mathbf{w}}_{j2}^*$

Compute the SDR solutions $\mathbf{W}_{j1}^*$ and $\mathbf{W}_{j2}^*$ by solving problem (57) with the fixed $\mathbf{Q}_k$, $\hat{a}_j$, $\mathbf{w}_{\bar{j}1}$, and $\mathbf{w}_{\bar{j}2}$.
for $i = 1$ to I **do**
  Generate samples $\hat{\mathbf{w}}_{j1} \sim \mathcal{CN}(0, \mathbf{W}_{j1}^*)$ and $\hat{\mathbf{w}}_{j2} \sim \mathcal{CN}(0, \mathbf{W}_{j2}^*)$;
  Set the second element of $\hat{\mathbf{w}}_{j2}$ as 1;
  if $\hat{\mathbf{w}}_{j1}^H \hat{\mathbf{w}}_{j1} + \left(\hat{\mathbf{w}_{j2}}^H \hat{\mathbf{w}}_{j2} - 1\right) m_j > P_j$ **then**
    update $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$ in (58) and (60).';
  end if
  if $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$ do not satisfy $\max(|\lambda_1|, |\lambda_2|) < 1$ **then**
    Drop the couple of vectors $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$.
  else
    Compute $\hat{R}_{sum}(i)$ with the corresponding $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$ by (61).
  end if
  $i = i + 1$.
end for
Search $i^* = \arg\max_{i \in \mathbf{C}} \hat{R}_{sum}(i)$ ($\mathbf{C} = \{1, \cdots, I\} - \{\text{dropped sequence}\}$);
Let $\hat{\mathbf{w}}_{j1}^*$ and $\hat{\mathbf{w}}_{j2}^*$ be the i^*-th corresponding vectors $\hat{\mathbf{w}}_{j1}$ and $\hat{\mathbf{w}}_{j2}$, respectively.

With the above setup, we reach the main algorithm of the Gaussian randomization method to optimize $\mathbf{w}_{j1}$ and $\mathbf{w}_{j2}$ with the fixed $\mathbf{w}_{\bar{j}1}$ and $\mathbf{w}_{\bar{j}2}$, which is summarized in Table 1.

After step 1), we update $\mathbf{Q}_k$ and $\hat{a}_j$ with the obtained transmission parameters $\{w_{jl}\}$ and then iteratively execute the two steps until the sum rate sequence R_{sum} converges to a finite constant. Here, R_{sum} is equivalent to $\hat{R}_{sum}$ in (61), whereas the matrices $\mathbf{T}_{\bar{j}kj}$ and $\mathbf{T}_{\bar{j}k\bar{j}}$ are updated by the obtained transmission parameters in Step 1). Based on Algorithm 1, the two-step iterative algorithm for solving problem (45) is summarized in Table 2. Here, without loss of generality, the initial transmission parameters of Algorithm 2 for each rate-profile α are obtained as follows: One transmitter transmits its own messages with the full power budget and the other one sorely forwards its counterpart's messages, i.e., $\mathbf{w}_{11}^0 = \left[0, \sqrt{P_1}\right]^T$, $\mathbf{w}_{12}^0 = [0, 1]^T$, $\mathbf{w}_{21}^0 = [0, 0]^T$, and $\mathbf{w}_{22}^0 = \left[\sqrt{\frac{P_2}{g_{21}P_1 + \hat{a}_2}}, 1\right]^T$.

TABLE 2. Algorithm 2: Two-step iterative algorithm for optimizing transmission parameters of problem (45) with a fixed rate-profile α.

Input: the power budget P_1 and P_2, the rate-profile α, two tolerances $\xi_0 > 0$ and $\xi_1 > 0$, $l = 0$, and the initial transmission parameters $\mathbf{w}_{11}^0$, $\mathbf{w}_{12}^0$, $\mathbf{w}_{21}^0$, and $\mathbf{w}_{22}^0$

Output: $\mathbf{w}_{11}^*$, $\mathbf{w}_{12}^*$, $\mathbf{w}_{21}^*$, $\mathbf{w}_{22}^*$

repeat
  Compute $\mathbf{Q}_k^l$ by (39) and (41), and $\hat{a}_j$ by Corollary 3.1 with $\mathbf{w}_{11}^l$, $\mathbf{w}_{12}^l$, $\mathbf{w}_{21}^l$, and $\mathbf{w}_{22}^l$;
  Let $(\mathbf{w}_{11}^{l+1}, \mathbf{w}_{12}^{l+1}, \mathbf{w}_{21}^{l+1}, \mathbf{w}_{22}^{l+1}) = (\mathbf{w}_{11}^l, \mathbf{w}_{12}^l, \mathbf{w}_{21}^l, \mathbf{w}_{22}^l)$.
  repeat
    Compute $(\mathbf{w}_{11}^{l+1}, \mathbf{w}_{12}^{l+1}) = (\hat{\mathbf{w}}_{11}^*, \hat{\mathbf{w}}_{12}^*)$ by Algorithm I with the fixed $\mathbf{w}_{21}^{l+1}$, $\mathbf{w}_{22}^{l+1}$, and $\mathbf{Q}_k^l$;
    Compute $(\mathbf{w}_{21}^{l+1}, \mathbf{w}_{22}^{l+1}) = (\hat{\mathbf{w}}_{21}^*, \hat{\mathbf{w}}_{22}^*)$ by Algorithm I with the fixed $\mathbf{w}_{11}^{l+1}$, $\mathbf{w}_{12}^{l+1}$, and $\mathbf{Q}_k^l$.
  until $|\hat{R}_{sum}(\mathbf{w}_{11}^{l+1}, \mathbf{w}_{12}^{l+1}) - \hat{R}_{sum}(\mathbf{w}_{21}^{l+1}, \mathbf{w}_{22}^{l+1})| < \xi_0$.
  $l = l + 1$.
until $|R_{sum}(\mathbf{w}_{11}^l, \mathbf{w}_{12}^l, \mathbf{w}_{21}^l, \mathbf{w}_{22}^l) - R_{sum}(\mathbf{w}_{11}^{l-1}, \mathbf{w}_{12}^{l-1}, \mathbf{w}_{21}^{l-1}, \mathbf{w}_{22}^{l-1})| < \xi_1$
Let $(\mathbf{w}_{11}^*, \mathbf{w}_{12}^*, \mathbf{w}_{21}^*, \mathbf{w}_{22}^*) = (\mathbf{w}_{11}^l, \mathbf{w}_{12}^l, \mathbf{w}_{21}^l, \mathbf{w}_{22}^l)$.

IV. SIMULATION RESULTS

In this section, we present some simulation results to validate our analysis, compared with the upper bound with source cooperation [14] and the capacity region of Gaussian IC with noncooperation [31]. In particular, the inner bounds of the noncooperation cases, with disabled transmitter cooperations, are specialized by utilizing two decoding schemes: joint decoding scheme and single-user decoding scheme further improved by optimal power control. To better demonstrate the cooperation gains, we consider a simple channel model [27]: the channel have fixed magnitude with i.i.d. random phase uniform between 0 and 2π. Hence, $h = |h|e^{j\theta}$ and $\theta \sim \mathbf{U}(0, 2\pi)$, which are fixed after channel realization. As a rule of thumb, we assume the cooperative channels $h_{12} = h_{21}$ and set the maximum power budget at each transmitter as $P_1 = P_2 = 20$ dB and the power of the CSCG noise as $\sigma_n^2 = 0$ dB. Moreover, for simplicity, we focus on a magnitude-symmetric case, i.e, $|h_{23}| = |h_{14}|$ and $|h_{13}| = |h_{24}|$. Note that all the simulation results are based on an average over 5000 channel phase realizations with the sample number $I = 100$ in Algorithm 1, and are all executed by CVX [32].

$$\mathcal{P}_0 = \left\{ \left(\mathbf{W}_{j1}, \mathbf{W}_{j2}\right) : \mathrm{Tr}(\mathbf{W}_{j1}) + \mathrm{Tr}(\mathbf{E}_1 \mathbf{W}_{j2}) m_j \le P_j,\ g_{\bar{j}\bar{j}} \mathrm{Tr}(\mathbf{E}_2 \mathbf{W}_{j1}) \le \bar{P}_{\bar{j}} \right\} \tag{56}$$

$$\hat{R}_{sum} \triangleq \min \left\{ \begin{array}{l} \frac{1}{\alpha} \log\left(\hat{\mathbf{w}}_{jj}^H \mathbf{T}_{\bar{j}3j} \hat{\mathbf{w}}_{jj} + 1\right), \frac{1}{\bar{\alpha}} \log\left(\hat{\mathbf{w}}_{\bar{j}\bar{j}}^H \mathbf{T}_{\bar{j}3\bar{j}} \hat{\mathbf{w}}_{\bar{j}\bar{j}} + 1\right), \\ \log\left(\hat{\mathbf{w}}_{\bar{j}\bar{j}}^H \mathbf{T}_{\bar{j}3\bar{j}} \hat{\mathbf{w}}_{\bar{j}\bar{j}} + \hat{\mathbf{w}}_{jj}^H \mathbf{T}_{\bar{j}3j} \hat{\mathbf{w}}_{jj} + 1\right), \frac{1}{\alpha} \log\left(\hat{\mathbf{w}}_{jj}^H \mathbf{T}_{\bar{j}4j} \hat{\mathbf{w}}_{jj} + 1\right), \\ \frac{1}{\bar{\alpha}} \log\left(\hat{\mathbf{w}}_{\bar{j}ji}^H \mathbf{T}_{\bar{j}4\bar{j}} \hat{\mathbf{w}}_{\bar{j}ji} + 1\right), \log\left(\hat{\mathbf{w}}_{\bar{j}\bar{j}}^H \mathbf{T}_{\bar{j}4\bar{j}} \hat{\mathbf{w}}_{\bar{j}\bar{j}} + \hat{\mathbf{w}}_{jj}^H \mathbf{T}_{\bar{j}4j} \hat{\mathbf{w}}_{jj} + 1\right) \end{array} \right\}, \tag{61}$$

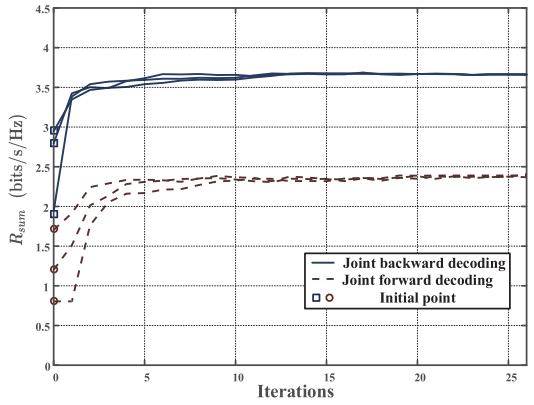


FIGURE 4. The convergency of Algorithm 2 for joint decoding.

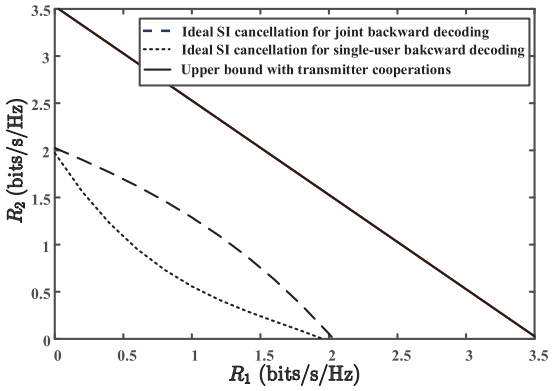


FIGURE 5. Achievable rate regions for ideal SI cancellation cases.

First, the convergency of Algorithm 2 is presented in Fig. 4 to see the variations of the object sum-rate R_{sum} as iterations increase. Typically, we consider the scenario for joint decoding scheme with the rate-profile $\alpha = 0.5$, i.e., $R_1 = R_2$, the residual SI power $\hat{P}_1 = \hat{P}_2 = 0$ dB, $h_{12} = h_{21} = 1$ and randomly setting $h_{13} = 0.0563 + 0.0826i$, $h_{24} = -0.0996 - 0.0084i$, $h_{14} = 0.3969 - 0.0499i$, and $h_{23} = -0.3986 - 0.0336i$. Then, the sum rate R_{sum} initiated from different initial transmission parameters over the iterations is plotted in Fig. 4, showing that as the number of iterations increases to approximately 13 or larger, the sum rates R_{sum} initiated at different points will converge to a same constant. Similar results can be obtained for single-user decoding scheme and most channel conditions.

Next, the AF achievable rate regions, considering the ideal SI cancellation, are plotted to compare with the capacity upper bound of the interference channel [14]. In particular, the forward decoding scheme is omitted here, due to its weakness in contrast to the backward decoding scheme (which is described latter). Hence, by setting $|h_{13}| = |h_{24}| = 0.1$, $|h_{23}| = |h_{14}| = 0.12$, and $|h_{12}| = |h_{21}| = 1$, as shown in Fig. 5, it is observed that the maximum sum rate of the joint backward decoding scheme can be achieved within around 1.2 bit of the upper bound under this scenario.

Fig. 6 and Fig. 7 plot the AF achievable rate regions with imperfect SI cancellation for joint and single-user decoding schemes to compare with the noncooperations cases.

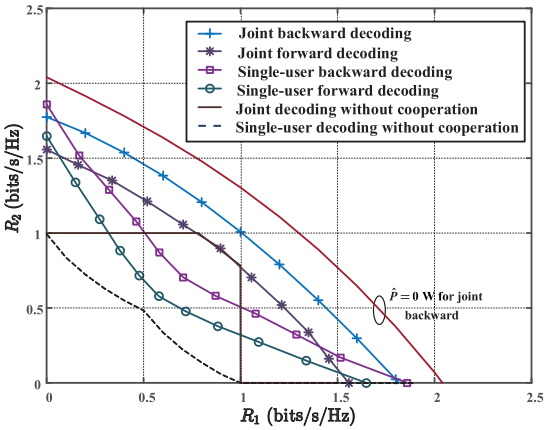


FIGURE 6. Achievable rate regions in a strong interference scenario ($|h_{13}| = |h_{24}| < |h_{23}| = |h_{14}|$): $|h_{13}| = |h_{24}| = 0.1$, $|h_{23}| = |h_{14}| = 0.12$ $|h_{12}| = |h_{21}| = 1$, and $\hat{P} = \hat{P}_1 = \hat{P}_2 = 0$ dB.

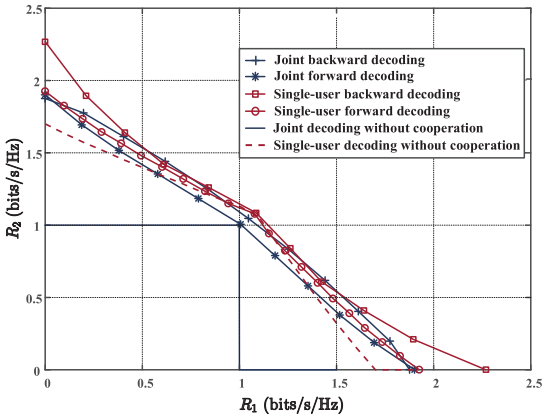


FIGURE 7. Achievable rate regions in a weak interference scenario ($|h_{13}| = |h_{24}| > |h_{23}| = |h_{14}|$): $|h_{13}| = |h_{24}| = 0.15$, $|h_{23}| = |h_{14}| = 0.1$ $|h_{12}| = |h_{21}| = 1$, and $\hat{P} = \hat{P}_1 = \hat{P}_2 = 0$ dB.

From Fig. 6, it is observed that despite decreasing around 0.5 bits/s/Hz from ideal SI cancellation case, the joint backward decoding scheme with residual SI still provides considerable rate gain compared with the noncooperation case, so do the other decoding schemes. For instance, the maximal sum rate gain is about 0.4 bits/s/Hz given by the joint backward decoding and about 0.7 bits/s/Hz given by the single-user backward decoding scheme. Furthermore, it is shown that in this case the joint decoding schemes always perform better than the single-user decoding schemes. It is because the strong interference, i.e., undesired signals are rather stronger than the desired ones. On the opposite, as shown in Fig. 7, the single-user decoding schemes perform better than the joint decoding schemes. It is due to the fact that in this case, the interference signals are much weaker than the desired signal since all the cross-channels are very weak. In addition, another important observation is that for all channel conditions, the backward decoding scheme is always better than forward decoding scheme, due to a intuition explanation that in the aforementioned two-slot decoding schemes, the potential power gain is dedicated to the backward time

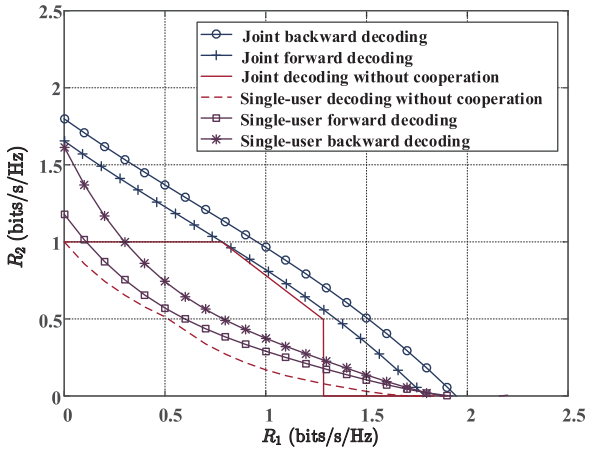


FIGURE 8. Achievable rate regions in an asymmetric scenario, i.e., $|h_{13}| \neq |h_{24}|$ and $|h_{23}| \neq |h_{14}|$.

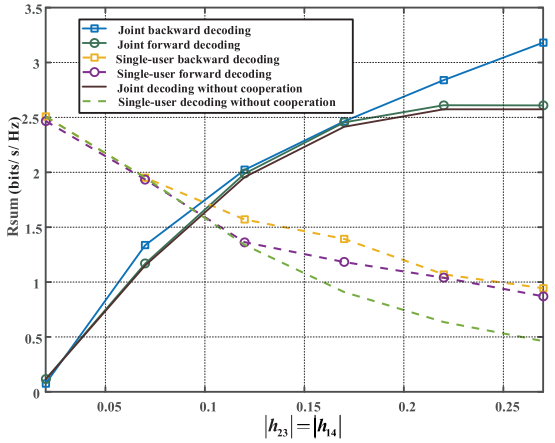


FIGURE 9. The effect of the cross-channel coefficients $|h_{23}| = |h_{14}|$ on the maximum R_{sum}.

slot, in which the forward decoding are not able to eliminate interferences induced by unknown time slot signals.

Fig. 8 plots the AF achievable rare regions in an asymmetric scenario where $|h_{13}| = 0.15$, $|h_{24}| = 0.1$, $|h_{23}| = 0.2$, $|h_{14}| = 0.1$, $|h_{12} = h_{21}| = 1$ and $\hat{P} = \hat{P}_1 = \hat{P}_2 = 0$ dB. It is observed that the proposed AF scheme also shows a good performance in the asymmetric case. To be details, the maximal sum rate of the joint backward decoding scheme is about 0.2 bits/s/Hz better than that of the joint non-cooperative scheme. In addition, the single-user backward decoding scheme performs better than the forward one and achieves about 0.3 bits/s/Hz better than the non-cooperation scheme.

Fig. 9 shows a plot of the maximum sum rate versus the cross channel coefficients $|h_{23}| = |h_{14}|$. Here, we set the rate-profile $\alpha = 0.5$, $|h_{12}| = |h_{21}| = 2$, $\hat{P}_1 = \hat{P}_2 = 0$ dB, and $|h_{13}| = |h_{24}| = 0.12$. From Fig. 9, it is observed that as the cross links become stronger, the rate improvement induced by the AF schemes will be enhanced. As $|h_{23}| = |h_{14}|$ increase beyond around 0.1, the joint decoding schemes are better than the single-user decoding schemes.

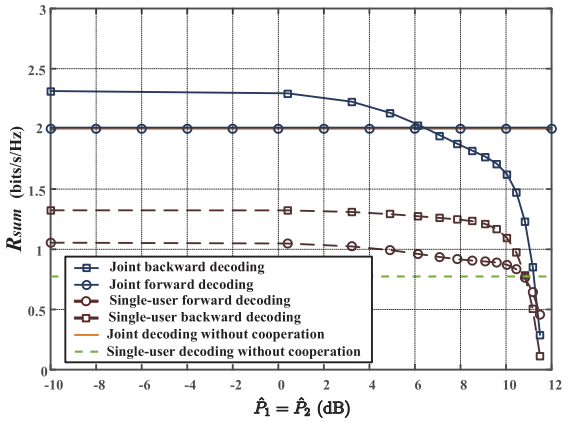


FIGURE 10. The effect of SI cancellation on the maximum R_{sum}.

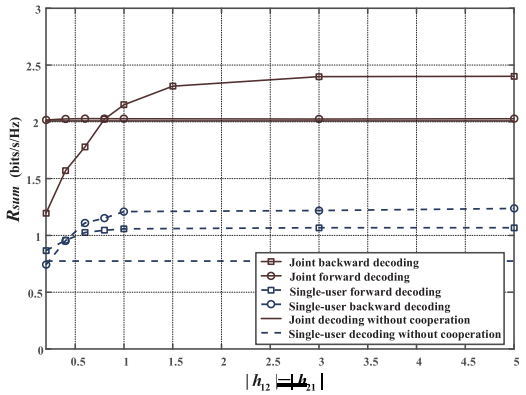


FIGURE 11. The effect of the cooperative channel coefficients $|h_{12}| = |h_{21}|$ on the maximum R_{sum}.

Fig. 10 studies the maximum sum rate R_{sum} versus the residual SI $\hat{P}_1 = \hat{P}_2$ by setting $\alpha = 0.5$, $|h_{12}| = |h_{21}| = 2$, $|h_{13}| = |h_{24}| = 0.1$, and $|h_{14}| = |h_{23}| = 0.15$. From Fig. 10, it is observed that the increasing of the residual SI power shows a negative impact on the performance of the system. For instance, as $\hat{P}_1 = \hat{P}_2$ increase to 8 dB or larger, the sum rate rapidly declines.

Fig. 11 plots the maximum sum rate R_{sum} versus the cooperative channel coefficients $|h_{12}| = |h_{21}|$. Here, we set $\hat{P}_1 = \hat{P}_2 = 0$ dB, $\alpha = 0.5$, $|h_{13}| = |h_{24}| = 0.1$, and $|h_{14}| = |h_{23}| = 0.15$, to show the effect of the cooperative channels on the achievable rates. In Fig. 11, as the channel coefficients $|h_{12}|$ and $|h_{21}|$ increase, we observe a significant increase in the rates. For instance, for the joint backward decoding scheme, the maximum rate R_{sum} increases by nearly 0.9 (bits/s/Hz) as $|h_{12}|$ and $|h_{21}|$ increase from 0.5 to 2. Note that in the small channel coefficients regime, the proposed schemes with transmitter cooperations perform worse than the cases without cooperation. The explanation is: When the cooperative channels go worse, as shown in Fig. 11, the performance benefits brought by the transmitter cooperations decrease and finally tend to zero. Hence, the optimal solution for the weak cooperative channels is to disable the transmitter cooperations. However, the proposed algorithm converges to a suboptimal solution still with the transmitter

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbb{E}\left[\left(\sum_{n=0}^{i-1}\mathbf{A}^n\mathbf{F}(i-n)\right)\left(\sum_{m=0}^{k-1}\mathbf{F}^H(k-m)(\mathbf{A}^m)^H\right)\right] \tag{64}$$

$$= \mathbb{E}\left[\left(\sum_{n=0}^{i-1}\mathbf{A}^n\mathbf{F}(i-n)\right)\left(\sum_{m=0}^{k-i-1}\mathbf{F}^H(k-m)(\mathbf{A}^m)^H + \sum_{m=k-i}^{k-1}\mathbf{F}^H(k-m)(\mathbf{A}^m)^H\right)\right] \tag{65}$$

$$= \mathbb{E}\left[\left(\sum_{n=0}^{i-1}\mathbf{A}^n\mathbf{F}(i-n)\right)\left(\begin{array}{c}\sum\limits_{m=0}^{k-i-1}\mathbf{F}^H(k-m)(\mathbf{A}^{m-k+i})^H \\ +\sum\limits_{m=0}^{i-1}\mathbf{F}^H(i-m)(\mathbf{A}^m)^H\end{array}\right)(\mathbf{A}^{k-i})^H\right] \tag{66}$$

$$= \left(\begin{array}{c}\sum\limits_{n=0}^{i-1}\sum\limits_{m=0}^{k-i-1}\mathbf{A}^n\mathbb{E}\left(\mathbf{F}(i-n)\mathbf{F}^H(k-m)\right)(\mathbf{A}^{m-k+i})^H \\ +\sum\limits_{n=0}^{i-1}\sum\limits_{m=0}^{i-1}\mathbf{A}^n\mathbb{E}\left(\mathbf{F}(i-n)\mathbf{F}^H(i-m)\right)(\mathbf{A}^m)^H\end{array}\right)\cdot(\mathbf{A}^{k-i})^H \tag{67}$$

$$= \left(\sum_{n=0}^{i-1}\sum_{m=0}^{i-1}\mathbf{A}^n\mathbb{E}\left(\mathbf{F}(i-n)\mathbf{F}^H(i-m)\right)(\mathbf{A}^m)^H\right)\cdot(\mathbf{A}^{k-i})^H \tag{68}$$

$$= \left(\sum_{n=0}^{i-1}\mathbf{A}^n\mathbb{E}\left(\mathbf{F}(i-n)\mathbf{F}^H(i-n)\right)(\mathbf{A}^n)^H\right)\cdot(\mathbf{A}^{k-i})^H \tag{69}$$

$$= \left(\sum_{n=0}^{i-1}\mathbf{P}\begin{bmatrix}\lambda_1^n & 0\\ 0 & \lambda_2^n\end{bmatrix}\mathbf{P}^{-1}\begin{bmatrix}\hat{P}_1+\sigma_n^2 & 0\\ 0 & \hat{P}_2+\sigma_n^2\end{bmatrix}\left(\mathbf{P}^{-1}\right)^H\begin{bmatrix}\lambda_1^{*n} & 0\\ 0 & \lambda_2^{*n}\end{bmatrix}\mathbf{P}^H\right)\cdot(\mathbf{A}^{k-i})^H \tag{70}$$

$$= \mathbf{P}\left(\sum_{n=0}^{i-1}\begin{bmatrix}|\lambda_1^n|^2 f_{11} & \lambda_1^n\lambda_2^{*n}f_{12}\\ \lambda_1^{*n}\lambda_2^n f_{12}^* & |\lambda_2^n|^2 f_{22}\end{bmatrix}\right)\begin{bmatrix}\lambda_1^* & 0\\ 0 & \lambda_2^*\end{bmatrix}^{k-i}\mathbf{P}^H \tag{71}$$

$$\to \mathbf{P}\begin{bmatrix}\frac{1}{1-|\lambda_1|^2}f_{11} & \frac{1}{1-\lambda_1\lambda_2^*}f_{12}\\ \frac{1}{1-\lambda_1^*\lambda_2}f_{12}^* & \frac{1}{1-|\lambda_2|^2}f_{22}\end{bmatrix}\begin{bmatrix}\lambda_1^* & 0\\ 0 & \lambda_2^*\end{bmatrix}^{k-i}\mathbf{P}^H. \tag{72}$$

cooperations and thus results in a worse performance than the noncooperative scheme.

V. CONCLUSION

In this paper, we proposed an AF transmission strategy via FD communications in a Gaussian IC. By analyzing the transmission process, we found an ARIN accumulation phenomenon between the two-transmitter cooperations, performing as a Markov process. Under a stationary state, the asymptotic achievable rates for the considered system were derived by both the joint and single-user decoding schemes. To characterize the achievable rate regions, an SDR-based iterative algorithm was proposed to optimize the transmission parameters. The results revealed that the proposed scheme achieves a better performance for all the channel conditions and the backward decoding performs better than the forward decoding.

APPENDIX A
PROOF OF PROPOSITION 2

To prove Proposition 2, we first compute the covariance matrix for the stochastic process $\{\mathbf{F}(i)\}$ in Proposition 1. Since the residual SI $\hat{s}_{j1}(i)$ and $n_j(i)$ are modeled as CSCG noise and are independent across different time slots and different transmitters, the covariance matrix of $\{\mathbf{F}(i)\}$ is obtained as

$$\mathbb{E}\left(\mathbf{F}(i)\mathbf{F}(i)^H\right) = \begin{bmatrix}\hat{P}_1+\sigma_n^2 & 0\\ 0 & \hat{P}_2+\sigma_n^2\end{bmatrix}, \tag{62}$$

$$\mathbb{E}\left(\mathbf{F}(i)\mathbf{F}(k)^H\right) = \mathbf{0}, \quad i\neq k, \tag{63}$$

where $\hat{P}_j$, $j=1,2$, is the power of the residual SI $\hat{s}_{j1}(i)$, and σ_n^2 is the power of the CSCG noise $n_j(i)$.

Based on Theorem 7.1.6 and Theorem 7.1.9 in [24], we can decompose matrix $\mathbf{A}$ as the following two cases:

- Case I: Matrix $\mathbf{A}$ has two different eigenvalues, i.e., $\lambda_1\neq\lambda_2$, and thus matrix $\mathbf{A}$ can be decomposed as (18) with a invertible matrix $\mathbf{P}$. If $k>i$, the covariance of $\hat{\mathbf{Y}}(i)$ can be calculated as (64)-(72), as shown at the top of this page.
  Here, we explain the above equations in details. From (17), (64) can be easily obtained. (66) can be obtained by extracting the common factor $(\mathbf{A}^{k-i})^H$ from (65). According to (62) and (63), it follows $\sum_{n=0}^{i-1}\sum_{m=0}^{k-i-1}\mathbf{A}^n\mathbb{E}\left(\mathbf{F}(i-n)\mathbf{F}^H(k-m)\right)(\mathbf{A}^{m-k+i})^H = \mathbf{0}$, and thus (68) and (69) can be obtained. By substituting

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbf{P}\left(\sum_{n=0}^{i-1}\begin{bmatrix}\lambda^n & n\lambda^{n-1}\\ 0 & \lambda^n\end{bmatrix}\begin{bmatrix}f_{11} & f_{12}\\ f_{12}^* & f_{22}\end{bmatrix}\begin{bmatrix}\lambda^{*n} & 0\\ n\lambda^{*n-1} & \lambda^{*n}\end{bmatrix}\right)\mathbf{P}^H(\mathbf{A}^{k-i})^H \quad (74)$$

$$= \mathbf{P}\left(\sum_{n=0}^{i-1}\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\right)\mathbf{P}^H(\mathbf{A}^{k-i})^H, \quad (75)$$

$$\to \mathbf{P}\begin{bmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{bmatrix}\begin{bmatrix}\lambda^* & 0\\ 1 & \lambda^*\end{bmatrix}^{k-i}\mathbf{P}^H, \quad (76)$$

(62) and (18) into (69), (70) can be easily obtained. By substituting (20) into (70), we can simplify the notations in (70) and the equivalent equation is given as (71). According to (71), when i tends to infinity, the sum of the matrices in (71) has each element converging to a finite constant if and only if there hold $|\lambda_1| < 1$, $|\lambda_2| < 1$, $|\lambda_1\lambda_2^*| < 1$, and $|\lambda_1^*\lambda_2| < 1$, which are equivalent to the constraint $\max(|\lambda_1|, |\lambda_2|) < 1$. Under this constraint, the covariance of $\hat{\mathbf{Y}}(i)$ converges to (72) as $i, k \to \infty$. Similarly, if $k \leq i$, it follows

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbf{P}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}^{i-k} \times \begin{bmatrix}\frac{1}{1-|\lambda_1|^2}f_{11} & \frac{1}{1-\lambda_1\lambda_2^*}f_{12}\\ \frac{1}{1-\lambda_1^*\lambda_2}f_{12}^* & \frac{1}{1-|\lambda_2|^2}f_{22}\end{bmatrix}\mathbf{P}^H. \quad (73)$$

Therefore, for Case I, (19) is satisfied.

- Case II: Matrix $\mathbf{A}$ has two identical eigenvalues, i.e., $\lambda_1 = \lambda_2 = \lambda$, and thus matrix $\mathbf{A}$ can be decomposed as (21) with an invertible matrix $\mathbf{P}$. If $k > i$, we obtain the covariance calculations for Case II similar as (64)-(69). By substituting (20) and (21) into (69), it follows (74)-(76), as shown at the top of this page.
where $a_{11} = |\lambda|^{2n}f_{11} + n\lambda^*|\lambda|^{2(n-1)}f_{12}^* + n\lambda|\lambda|^{2(n-1)}f_{12} + n^2|\lambda|^{2(n-1)}f_{22}$, $a_{21}^* = a_{12} = |\lambda|^{2n}f_{12} + n\lambda^*|\lambda|^{2(n-1)}f_{22}$, and $a_{22} = |\lambda|^{2n}f_{22}$. Under the constraint $|\lambda| < 1$ and $i, k \to \infty$, (75) converges to (76) with $\{c_{ij}\}$ being defined in (23)-(25). Similarly, if $k \leq i$, it follows

$$\mathbb{E}\left(\hat{\mathbf{Y}}(i)\hat{\mathbf{Y}}(k)^H\right) = \mathbf{P}\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}^{i-k} \times \begin{bmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{bmatrix}\mathbf{P}^H. \quad (77)$$

Hence, (22) is satisfied for Case II.

With the above analysis, Proposition 2 can be proved.

ACKNOWLEDGMENT

This paper was presented in part at the IEEE SPAWC 2018.

REFERENCES

[1] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.

[2] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math., Statist., Probab.*, 1961, vol. 1, no. 3, pp. 611–644.

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

[4] 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.

[5] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, Jun. 2004.

[6] R. K. Farsani, "On the capacity region of the two-user interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 2724–2738.

[7] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 5, pp. 569–570, Sep. 1975.

[8] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.

[9] I. Maric, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.

[10] P. K. Sharma and P. Garg, "Achieving high data rates through full duplex relaying in multicell environments," *Trans. Emerg. Telecommun. Technol.*, vol. 27, no. 1, pp. 111–121, Oct. 2014.

[11] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2941–2965, May 2011.

[12] H. Bagheri, A. S. Motahari, and A. K. Khandani, "Zero-forcing for the symmetric interference channel with conferencing encoders," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010, pp. 370–374.

[13] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.

[14] V. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 156–186, Jan. 2011.

[15] Q. Li, K. H. Li, and K. C. Teh, "DMT of interference channels with full-duplex source cooperation and imperfect CSIT," in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, Dec. 2010, pp. 1–5.

[16] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, May 1983.

[17] P. K. Sharma and P. Garg, "Performance analysis of full duplex decode-and-forward cooperative relaying over Nakagami-*m* fading channels," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 9, pp. 905–913, Oct. 2013.

[18] P. K. Sharma and P. Garg, "Outage analysis of full duplex decode and forward relaying over Nakagami-*m* channels," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2013, pp. 1–5.

[19] J. Huang, D. Wang, and C. Huang, "Full-duplex amplify-and-forward transmitter cooperations for compound multiple access channels," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[20] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

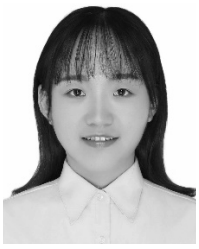
[21] H. Hamazumi, K. Imamura, N. Iai, K. Shibuya, and M. Sasaki, "A study of a loop interference canceller for the relay stations in an SFN for digital terrestrial broadcasting," in *Proc. IEEE Global Telecommun. Conf. Rec. (Globecom)*, Nov./Dec. 2000, pp. 167–171.

[22] C. Gardiner, *Stochastic Methods*. Berlin, Germany: Springer, 2009.

[23] S. Galli, "A simple two-tap statistical model for the power line channel," in *Proc. IEEE Int. Symp. Power Line Commun. (ISPLC)*, Mar. 2010, pp. 242–248.

[24] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.
[25] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.
[26] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[27] C. T. K. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, "Capacity gain from two-transmitter and two-receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3822–3827, Oct. 2007.
[28] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.
[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[30] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
[31] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[32] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," Dec. 2011, *arXiv:1001.3404*. [Online]. Available: https://arxiv.org/abs/1001.3404

JIANHAO HUANG received the B.S. degree in electrical engineering from Harbin Engineering University, Harbin, China, in 2017. He has been a Research Assistant with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China, since 2017. He is currently a Research Assistant with the School of Electrical Engineering and Intelligentization, Dongguan University of Technology. His current research interests include compress sensing in communication systems and full-duplex cooperation communications. He had been a TPC Member in GLOBECOM 2019. He is also a Reviewer of the IEEE Wireless Communications Letters.



DAN WANG received the B.S. degree in electrical engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2017. She is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include game theory in communication systems and full-duplex communications. She had been a TPC Member in GLOBECOM 2019. She is currently a Reviewer of the IEEE Wireless Communications Letters.

CHUAN HUANG (S'09–M'13) received the B.S. degree in mathematics and the M.S. degree in communications engineering from the University of Electronic Science and Technology of China, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2012. From August 2012 to December 2013, he was a Postdoctoral Research Fellow and an Assistant Research Professor, from December 2013 to July 2014, both at Arizona State University, Tempe, AZ, USA. He was a Visiting Scholar with the National University of Singapore and a Research Associate with Princeton University, respectively. He is currently a Professor with the School of Electrical Engineering and Intelligentization, Dongguan University of Technology. His current research interests include wireless communications and signal processing. He currently serves as an Editor for the IEEE Transactions on Wireless Communications, the IEEE Access, and the IEEE Wireless Communications Letters.