IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

# Codewords With Memory Improve Achievable Rate Regions of the Memoryless Gaussian Interference Channel

**Wasim Huleihel and Neri Merhav**

**CCIT Report #886**
**September 2015**

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #886 September 2015



# Codewords With Memory Improve Achievable Rate Regions of the Memoryless Gaussian Interference Channel

Wasim Huleihel and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 3200003, ISRAEL
E-mail: {wh@campus, merhav@ee}.technion.ac.il

**Abstract**

The two-user Gaussian interference channel (GIC) has been extensively studied in the literature during the last four decades. The full characterization of the capacity region of the GIC is a long-standing open problem, except the case of strong or very strong interference. For general GIC's, many inner bounds have been provided over the years, among of them, the Han-Kobayashi (HK) region, is the most celebrated one. Unfortunately, the calculation of the HK region is prohibitively complex, due to the appearance of some auxiliary random variables, whose optimal choice is an open problem. As in other multi-user communication systems, these achievable regions are based on ensembles of i.i.d. (memoryless) codewords, in the sense that the symbols within each codeword are drawn independently. In this paper, we show that for the GIC, it is worthwhile to employ random coding ensembles of codewords with memory. Specifically, we take known achievable regions for the GIC, and generalize/improve them by allowing dependency between the code symbols. For example, we improve the state-of-the-art HK region by drawing the codewords (of each codeword and for each user) from a first-order autoregressive moving average (ARMA) Gaussian process. In this way, we suggest several new achievable rate regions, which are easily calculable, and which are strictly better than state-of-the-art known achievable regions.

**Index Terms**

*This research was partially supported by The Israeli Science Foundation (ISF), grant no. 412/12.

Interference channel (IC), achievable rate region, bounds, capacity region, Han-Kobayashi region.

## I. Introduction

The two-user interference channel (IC) models a general scenario of communication between two transmitters and two receivers (with no cooperation at either side), where each receiver decodes its intended message from an observed signal, which is interfered by the other user, and corrupted by channel noise. The information-theoretic analysis of this model has begun over more than four decades ago and has recently witnessed a resurgence of interest. In this paper, we focus on the two-user Gaussian interference channel (GIC), which has been extensively studied in the literature, see, for example, [1, Ch. 6], [2], and many references therein. In the following, we describe briefly the two-user GIC model in its *standard form*, adopting to the notation in [3]. The discrete-time memoryless GIC is defined as follows:

$$Y_1 = X_1 + \sqrt{a_{12}} X_2 + Z_1, \tag{1a}$$

$$Y_2 = \sqrt{a_{21}} X_1 + X_2 + Z_2, \tag{1b}$$

where $(X_1, X_2)$ and $(Y_1, Y_2)$ the real-valued inputs and outputs, respectively, the cross-link gains $a_{12}$ and $a_{21}$ are assumed time-invariant, and $Z_1$ and $Z_2$ are Gaussian random variables with zero mean and unit variance. We let $X_1^n \triangleq (X_{1,1}, X_{1,2}, \ldots, X_{1,n})$ and $X_2^n \triangleq (X_{2,1}, X_{2,2}, \ldots, X_{2,n})$ be two transmitted codewords across the channel, for each user, where $X_{i,j}$ designates the symbol that is transmitted by user $i \in \{1, 2\}$ at time instant $j \in \{1, 2, \ldots, n\}$. As usual, we assume that the receivers have full knowledge of the codebooks used by both users, and that there is no cooperation between the transmitters, and between the receivers. The following power constraints are imposed on the transmitted signals:

$$\sum_{j=1}^{n} \mathbb{E}(X_{i,j}^2) \le nP_i \tag{2}$$

for $i = 1, 2$, where $P_1, P_2 > 0$. It is assumed that the random vectors $Z_1^n$ and $Z_2^n$ have i.i.d. Gaussian entries with zero mean and unit variance, and they are independent of the inputs. Without loss of generality, we assume that $Z_1^n$ and $Z_2^n$ are independent, due to the fact that the capacity region of a two-user, discrete-time, memoryless IC depends only on the marginal probability density functions (pdfs) $p(y_i|x_1, x_2)$ for $i \in \{1, 2\}$.

Despite of its simplicity, the problem of specifying a computable expression of the capacity region for the two-user GIC is still open, although it has been solved for some special cases of strong interference [4, 5], where $a_{12}, a_{21} \ge 0$, or very strong interference [6] where $a_{12} \ge 1 + P_1$ and $a_{21} \ge 1 + P_2$. Recently, however, the corner points of the capacity region were fully determined [7, 8], thus approving

Costa's conjecture [9, 10]. The study of the memoryless IC started in [11], and continued in [12], where the capacity region of a general discrete memoryless IC was expressed by a limiting expression (multi-letter formula) which, unfortunately, does not lend itself to a computable expression. Fundamental simple inner and outer bounds to the capacity region were also determined. In [13] an improved achievable rate region was derived, by using the well-known superposition coding technique. Various inner and outer bounds were established in [14] by transforming the original problem to some associated multiple access or broadcast channel settings. In their important work, Han and Kobayashi (HK) [4] have derived an inner bound for a general discrete memoryless IC. It includes the achievable regions that were earlier established, and it is still the best known inner bound. Unfortunately, the computation of the HK achievable region is prohibitively complex, due to the appearance of some auxiliary random variables, whose optimal choice is an open problem. A simplified description of the HK region for the IC was derived in [15]. Finally, in [3], another achievable region was derived, based on a modified time (or frequency) division multiplexing approach, originated in [14]. The suggested region is easily calculable, though it is a special case of the celebrated HK region. For a comprehensive survey paper on the IC, see [16].

Recently, in [17] the sub-optimality of the HK region was demonstrated numerically on the clean Z-interference (ZIC) channel with binary inputs and outputs. In Section 2.2 of the same paper, the authors provided some intuition, based on their findings, why i.i.d. coding (in the sense of HK) might not be optimal for the IC. Their intuition was based on the observation that $X_2$ acts like a state variable on the communication of the channel between $X_1$ and $Y_1$. Now if the channel from $X_1$ to $Y_1$, with $X_2$ as the state, is not memoryless, we know that the optimal code distributions $X_1^n$ are not independent distributions. Moreover, for the GIC, it is reasonable to believe that if the users will transmit on different frequency bands (or, for example, employ orthogonal signals), then there will be an improvement in the performance. These intuitions raise the following question:

**Question** *Can we improve known achievable rate regions by allowing dependency between the components, within each codeword of each user, rather than using memoryless (i.i.d.) distributions?*

The problem is that when the channel is discrete (in amplitude), it is extremely difficult to answer these questions. For example, assume that the codewords are drawn uniformly from a given Markov type. Then, analyzing the probability of error (associated for example with the joint typicality decoder) would require the calculation of entropy rates of certain hidden Markov processes, which is a well-known open problem on its own right. Fortunately, for GIC's, this task is easier, because asymptotic mutual information rates have compact expressions in terms of the relevant spectra.

Motivated by these observations, in this paper, we derive several new calculable achievable rate regions (some of them are based on known achievable regions), where contrary to all other previous works, we allow dependency between the components within each codeword. Specifically, we start with a general coding theorem, which modifies the region proposed in [3], by employing codewords sampled from processes of a general spectral density (subject to the power constraint), namely, the codewords are drawn from a general stationary Gaussian process, rather than standard i.i.d. random coding. Then, we consider a simple special case of a first-order[1] autoregressive moving average (ARMA) process. We show that this choice can significantly improve on the region proposed in [3], and in particular, on its corner points. Then, we consider more sophisticated encoding/decoding schemes that are based on the HK scheme, and propose two new simple and calculable inner bounds that considerably improve the aforementioned regions, and improve state-of-the-art sub-regions deduced from the HK inner bound.

The paper is organized as follows. In Section II, we formalize the problem and assert the main theorems. Specifically, in Subsection II-A, we modify Sato's and Sason's schemes by employing codewords of a general spectral density. Then, to demonstrate the improvement in using such schemes, we analyze the case of first-order ARMA process. We provide some numerical examples, and show that the new rate regions significantly improve on Sato's and Sason's results. Then, in Subsection II-B, we consider schemes that are based on the HK scheme, and derive new achievable regions. As before, we provide numerical examples which demonstrate the improvement compared to other known results. Finally, Section III is devoted to our main conclusions, and we also outline some possible extensions.

## II. Main Results

In this section, we present our main results and then discuss them. We first establish some notation conventions.

In our model, each sender, $k \in \{1,2\}$, wishes to communicate an independent message $M_k \in \left\{1,2,\ldots,2^{nR_k}\right\}$ at rate $R_k$, and each receiver, $l \in \{1,2\}$, wishes to decode its respective message. Specifically, a $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}_n$ consists of:

- Two message sets $\mathcal{M}_1 \triangleq \left\{0,\ldots,2^{nR_1}-1\right\}$ and $\mathcal{M}_2 \triangleq \left\{0,\ldots,2^{nR_2}-1\right\}$ for the first and second users, respectively.
- Two encoders, where for each $k \in \{1,2\}$, the $k$-th encoder assigns a codeword $\boldsymbol{x}_{k,i}$ to each message $i \in \mathcal{M}_k$.
- Two decoders, where each decoder $l \in \{1,2\}$ assigns an estimate $\hat{M}_l$ to $M_l$.

[1]By "first-order", we mean one AR parameter and one MA parameter (i.e., one pole and one zero).

We assume that the message pair $(M_1, M_2)$ is uniformly distributed over $\mathcal{M}_1 \times \mathcal{M}_2$. For a coded information of block length $n$, the two-user memoryless IC is denoted by:

$$(\mathcal{X}_1^n \times \mathcal{X}_2^n, P^n(y_1^n, y_2^n | x_1^n, x_2^n), \mathcal{Y}_1^n \times \mathcal{Y}_2^n) \tag{3}$$

where

$$P^n(y_1^n, y_2^n | x_1^n, x_2^n) = \prod_{k=1}^{n} P(y_{1,k}, y_{2,k} | x_{1,k}, x_{2,k}), \tag{4}$$

and in our GIC case $P(y_{1,k}, y_{2,k} | x_{1,k}, x_{2,k})$ is derived from (1). Since there is no cooperation between the two receivers, the average probabilities of error are

$$P_{e,i}^{(n)} = 2^{-n(R_1+R_2)} \sum_{m_1,m_2} \mathbb{P}\left\{ \hat{M}_i\left(Y_i^n\right) \neq m_i | M_1 = m_1, M_2 = m_2 \right\}, \quad i = 1, 2. \tag{5}$$

A rate pair $(R_1, R_2)$ is said to be achievable if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes, such that $P_{e,1}^{(n)} \to 0$ and $P_{e,2}^{(n)} \to 0$, as $n \to \infty$. The rates are expressed here in terms of bits per channel use. The capacity region of an IC is defined as the closure of the set of all its achievable rate pairs.

As mentioned in the introduction, previous works on the IC focused mainly on standard random coding, where each codeword is independently and identically generated according to some given probability distribution. However, motivated by the above-mentioned interesting observation of [18], it is desirable to analyze other ensembles with possible dependency between the components of each codeword.

In Subsection II-A below, we derive a new achievable region which is based on Sato's and Sason's schemes [3, 14]. Then, in Subsection II-B, we derive inner bounds which are based on the HK scheme [4].

### *A. Sato-Sason-based inner bound:*

In [14, Theorem 5], it was shown that the following rate-region is achievable:

$$\mathcal{G}_B = \text{conv}\left\{\mathcal{G}_{B_1} \cup \mathcal{G}_{B_2}\right\} \tag{6}$$

where conv$\{\cdot\}$ denotes the convex closure of a set, and where

$$\mathcal{G}_{B_1} \triangleq \text{conv} \bigcup_{P_{X_1}, P_{X_2}} \left\{ (R_1, R_2) : \begin{array}{c} 0 \leq R_1 \leq I(X_1; Y_1 | X_2) \\ 0 \leq R_2 \leq \min\left\{ I(X_2; Y_1), I(X_2; Y_2) \right\} \end{array} \right\} \tag{7}$$

and

$$\mathcal{G}_{B_2} \triangleq \text{conv} \bigcup_{P_{X_1}, P_{X_2}} \left\{ (R_1, R_2) : \begin{array}{c} 0 \leq R_1 \leq \min\left\{ I(X_1; Y_1), I(X_1; Y_2) \right\} \\ 0 \leq R_2 \leq I(X_2; Y_2 | X_1) \end{array} \right\}. \tag{8}$$

This achievable region is interpreted as follows: the first region, $\mathcal{G}_{B_1}$, refers to the mode of work where receiver 1 first decodes the message of the second sender, and then uses it as side information for decoding his own message. Receiver 2 directly decodes his message based on his received signal. The second region, $\mathcal{G}_{B_2}$, refers to the opposite mode of work, where the roles of receiver 1 and 2 are interchanged. Sason's scheme [3] is based on the above region. His idea was to employ the first mode for a fraction $\lambda$ of the transmission time, and the second mode for the complementary time. Let us define generically, $\bar{z} \triangleq 1-z$ for $z \in [0,1]$, and $\eta(x) \triangleq 0.5\log_2(1+x)$ for $x \geq 0$. The following theorem states an achievability result.

*Theorem 1 ([3, Th. 1])* The set of rate pairs:

$$\mathcal{R}_{\text{sason}} = \bigcup_{\alpha,\beta,\lambda\in[0,1]} \left\{ (R_1,R_2) : \begin{array}{l} R_1 \leq \lambda \cdot \eta\left(\frac{\alpha P_1}{\lambda}\right) + \bar{\lambda} \cdot \min\left\{\eta\left(\frac{\bar{\alpha} P_1}{\lambda + a_{12}\bar{\beta}P_2}\right), \eta\left(\frac{a_{21}\bar{\alpha}P_1}{\lambda+\beta P_2}\right)\right\} \\ R_2 \leq \bar{\lambda}\cdot\eta\left(\frac{\bar{\beta}P_2}{\bar{\lambda}}\right) + \lambda \cdot \min\left\{\eta\left(\frac{\beta P_2}{\lambda + a_{21}\alpha P_1}\right), \eta\left(\frac{a_{12}\beta P_2}{\lambda+\alpha P_1}\right)\right\} \end{array} \right\} \tag{9}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

In order to obtain the above region, Sason [14] employed a random coding approach, in which the codewords are selected in a *memoryless* manner, that is, they are drawn independently from a given probability distribution. Since in our coding scheme there will be dependency between the components of each codeword, the above single-letter expression will no longer be suitable. Let $\{r_{11,i}\}_{i=0}^{\infty}, \{r_{12,i}\}_{i=0}^{\infty}, \{r_{21,i}\}_{i=0}^{\infty}$ and $\{r_{22,i}\}_{i=0}^{\infty}$ be square summable, non-negative definite sequences[2] with $r_{11,0} = r_{12,0} = r_{21,0} = r_{22,0} = 1$. Let $\alpha,\beta,\lambda \in [0,1]$, and define:

$$\nu_{11}(\omega) \triangleq \frac{\alpha P_1}{\lambda}\left(1 + 2\sum_{k=1}^{\infty} r_{11,k}\cos(k\omega)\right), \tag{10a}$$

$$\nu_{21}(\omega) \triangleq \frac{\beta P_2}{\lambda}\left(1 + 2\sum_{k=1}^{\infty} r_{21,k}\cos(k\omega)\right), \tag{10b}$$

$$\nu_{12}(\omega) \triangleq \frac{\bar{\alpha} P_1}{\bar{\lambda}}\left(1 + 2\sum_{k=1}^{\infty} r_{12,k}\cos(k\omega)\right), \tag{10c}$$

$$\nu_{22}(\omega) \triangleq \frac{\bar{\beta} P_2}{\bar{\lambda}}\left(1 + 2\sum_{k=1}^{\infty} r_{22,k}\cos(k\omega)\right), \tag{10d}$$

where $\omega \in [0, 2\pi)$. Finally, we define the functional:

$$\varphi[f] \triangleq \frac{1}{4\pi}\int_0^{2\pi} \log_2[1+f(\omega)]\,\mathrm{d}\omega. \tag{11}$$

[2] A sequence $\{a_l\}_{l=-\infty}^{\infty}$ is called non-negative definite if for any $n \geq 1$, the Toeplitz matrix, $\boldsymbol{A}^{n\times n}$, that is generated by $\{a_l\}_{l=-(n-1)}^{n-1}$, is non-negative definite. Equivalently, $\{a_l\}_{l=-\infty}^{\infty}$ is non-negative definite if the spectral density $\sum_{k=-\infty}^{\infty} a_k \exp(jk\omega)$ is non-negative for any $\omega \in [0,2\pi]$. If $a_k = a_{-k}$ for $k \geq 1$, then the sequence $\{a_l\}_{l=-\infty}^{\infty}$ is non-negative definite if $\sum_{k=0}^{\infty} a_k \cos(k\omega)$ is non-negative for any $\omega \in [0,2\pi]$.

Let $\boldsymbol{\theta} \triangleq (\alpha, \beta, \lambda, \{r_{11,i}\}_{i\geq 1}, \{r_{12,i}\}_{i\geq 1}, \{r_{21,i}\}_{i\geq 1}, \{r_{22,i}\}_{i\geq 1}) \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the support of $\boldsymbol{\theta}$, as defined above. The following theorem is proved in Appendix A.

*Theorem 2* The set of rate pairs:

$$\mathcal{R} = \bigcup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq \lambda \cdot \varphi\left[\nu_{11}\right] + \bar{\lambda} \cdot \min\left\{ \varphi\left[\frac{\nu_{12}}{1+a_{12}\nu_{22}}\right], \varphi\left[\frac{a_{21}\nu_{12}}{1+\nu_{22}}\right] \right\} \\ R_2 \leq \bar{\lambda} \cdot \varphi\left[\nu_{22}\right] + \lambda \cdot \min\left\{ \varphi\left[\frac{a_{12}\nu_{21}}{1+\nu_{11}}\right], \varphi\left[\frac{\nu_{21}}{1+a_{21}\nu_{11}}\right] \right\} \end{array} \right\}, \tag{12}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

Note that Theorem 2 is a generalization of Theorem 1. Indeed, in order to obtain Theorem 1, Sason employed the encoding and decoding schemes mentioned above (see discussion before Theorem 1), where during a fraction $\lambda$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are Gaussian with zero mean, and variances $\alpha P_1/\lambda$ and $\beta P_2/\lambda$, respectively. During the remaining fraction $\bar{\lambda}$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are Gaussian with zero mean, and variances $\bar{\alpha} P_1/\bar{\lambda}$ and $\bar{\beta} P_2/\bar{\lambda}$, respectively. To obtain Theorem 2, on the other hand, we employ the same encoding and decoding schemes, but we assume that the codewords (for each user and each mode), are drawn from stationary Gaussian processes, with zero mean, and Toeplitz covariance matrices with the elements $\{r_{11,i}\}, \{r_{12,i}\}, \{r_{21,i}\}$ and $\{r_{22,i}\}$, where $r_{11,i} = (\alpha P_1/\lambda)^{-1}\mathbb{E}\{X_{1,m}X_{1,m+i}\}$, $r_{21,i} = (\beta P_2/\lambda)^{-1}\mathbb{E}\{X_{2,m}X_{2,m+i}\}$, for $1 \leq m+i \leq n\lambda - 1$, and $r_{12,i} = (\bar{\alpha} P_1/\bar{\lambda})^{-1}\mathbb{E}\{X_{1,m+n\lambda}X_{1,m+n\lambda+i}\}$, and $r_{22,i} = (\bar{\beta} P_2/\bar{\lambda})^{-1}\mathbb{E}\{X_{2,m+n\lambda}X_{2,m+n\lambda+i}\}$.

Theorem 2 is, of course, general and it requires a union over all combinations of user spectra (that satisfy the power constraint), so it is not calculable practically. To demonstrate the improvement of using codewords with memory, we propose the following special case (the complete details can be found in Appendix B). Consider the case where during a fraction $\lambda$ of the transmission time, the components of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by:

$$x_{1,i} = \rho_{x_1} x_{1,i-1} + \sigma_1 w_{1,i} - \kappa_1 \sigma_1 w_{1,i-1}, \;\; i = 1, 2, \ldots, n\lambda - 1 \tag{13a}$$

$$x_{2,i} = \rho_{x_2} x_{2,i-1} + \sigma_2 w_{2,i} - \kappa_2 \sigma_2 w_{2,i-1}, \;\; i = 1, 2, \ldots, n\lambda - 1 \tag{13b}$$

where, without essential loss of generality, we assume that $n\lambda$ is integer, and

$$\sigma_k^2 \triangleq \frac{1 - \rho_{x_k}^2}{1 + \kappa_k^2 - 2\kappa_k \rho_{x_k}}, \tag{14}$$

for $k = 1, 2$ where $|\rho_{x_1}|, |\rho_{x_2}|, |\kappa_1|, |\kappa_2| < 1$, $x_{1,0}$ and $x_{2,0}$ are Gaussian random variables with zero means and variances $\alpha P_1/\lambda$ and $\beta P_2/\lambda$, $\{w_{1,i}\}$ and $\{w_{2,i}\}$ are i.i.d. Gaussian process with zero mean and variances $\alpha P_1/\lambda$ and $\beta P_2/\lambda$, and $0 \leq \alpha, \beta \leq 1$. That is, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are distributed according to a *first-order auto-regressive moving average* (ARMA) model. Note that when $\kappa_1 = \kappa_2 = 0$, we have a *first-order Gaussian Markov process* (AR process). During the remaining fraction $\bar{\lambda} = 1 - \lambda$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by:

$$x_{1,i} = \rho_{x_1} x_{1,i-1} + \sigma_1 \tilde{w}_{1,i} - \kappa_1 \sigma_1 \tilde{w}_{1,i-1}, \;\; i = n\lambda + 1, \ldots, n \tag{15a}$$

$$x_{2,i} = \rho_{x_2} x_{2,i-1} + \sigma_2 \tilde{w}_{2,i} - \kappa_2 \sigma_2 \tilde{w}_{2,i-1}, \;\; i = n\lambda + 1, \ldots, n \tag{15b}$$

where $x_{1,n\lambda}$ and $x_{1,n\lambda}$ are Gaussian random variables with zero means and variances $\bar{\alpha}P_1/\bar{\lambda}$ and $\bar{\beta}P_2/\bar{\lambda}$, respectively, $\tilde{w}_{1,i}$ and $\tilde{w}_{2,i}$ are i.i.d. Gaussian process with zero means and variances $\bar{\alpha}P_1/\bar{\lambda}$ and $\bar{\beta}P_2/\bar{\lambda}$. When $\kappa_1 = \kappa_2 = 0$, it is evident that for $i = 1, \ldots, n\lambda - 1$, consecutive samples from $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have correlations $\rho_{x1}$ and $\rho_{x2}$, respectively. This is also the case for $i = n\lambda, \ldots, n$. The receiver, on the other hand, has the same two modes of work, as described earlier. Let:

$$\gamma_1^{(1)}(\omega) \triangleq \frac{\sigma_1^2 \alpha P_1 \left|1 - \kappa_1 e^{j\omega}\right|^2}{\lambda \left|1 - \rho_{x_1} e^{j\omega}\right|^2}, \tag{16a}$$

$$\gamma_2^{(1)}(\omega) \triangleq \frac{\sigma_2^2 \beta P_2 \left|1 - \kappa_2 e^{j\omega}\right|^2}{\lambda \left|1 - \rho_{x_2} e^{j\omega}\right|^2}, \tag{16b}$$

$$\gamma_1^{(2)}(\omega) \triangleq \frac{\sigma_1^2 \bar{\alpha} P_1 \left|1 - \kappa_1 e^{j\omega}\right|^2}{\bar{\lambda} \left|1 - \rho_{x_1} e^{j\omega}\right|^2}, \tag{16c}$$

$$\gamma_2^{(2)}(\omega) \triangleq \frac{\sigma_2^2 \bar{\beta} P_2 \left|1 - \kappa_2 e^{j\omega}\right|^2}{\bar{\lambda} \left|1 - \rho_{x_2} e^{j\omega}\right|^2}, \tag{16d}$$

where $\omega \in [0, 2\pi)$, $j \triangleq \sqrt{-1}$, and define $\boldsymbol{\psi} \triangleq (\alpha, \beta, \lambda, \rho_{x_1}, \rho_{x_2}, \kappa_1, \kappa_2)$, and $\boldsymbol{\Psi} \triangleq [0,1]^3 \times (-1,1)^4$. We have the following result, proved in Appendix B.

*Theorem 3* The set of rate pairs:

$$\mathcal{R}^{(1)} = \bigcup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq \lambda \cdot \varphi\left[\gamma_1^{(1)}\right] + \bar{\lambda} \cdot \min\left\{\varphi\left[\frac{\gamma_1^{(2)}}{1 + a_{12}\gamma_2^{(2)}}\right], \varphi\left[\frac{a_{21}\gamma_1^{(2)}}{1 + \gamma_2^{(2)}}\right]\right\} \\ R_2 \leq \bar{\lambda} \cdot \varphi\left[\gamma_2^{(2)}\right] + \lambda \cdot \min\left\{\varphi\left[\frac{a_{12}\gamma_2^{(1)}}{1 + \gamma_1^{(1)}}\right], \varphi\left[\frac{\gamma_2^{(1)}}{1 + a_{21}\gamma_1^{(1)}}\right]\right\} \end{array} \right\} \tag{17}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

Several remarks regarding Theorem 3 are in order:

- The above analysis can be generalized to an ARMA process of any order, i.e.,

$$x_{1,i} = \sum_{l=1}^{p_1} \rho_{l,x_1} x_{1,i-l} + \sigma_1 w_{1,i} - \sigma_1 \cdot \sum_{l=1}^{q_1} \kappa_{1,l} w_{1,i-l} \tag{18}$$

where the variance $\sigma_1^2$ should be chosen such that the $\text{Var}\left\{x_{1,i}\right\} = \alpha P_1/\lambda$ for all $1 \leq i \leq \lambda n$, and similarly for the second user. Then, we obtain the same result as before, but with:

$$\gamma_1^{(1)}(\omega) \triangleq \frac{\sigma_1^2 \alpha P_1 \left|1 - \sum_{l=1}^{q_1} \kappa_{1,l} e^{jl\omega}\right|^2}{\lambda \left|1 - \sum_{l=1}^{p_1} \rho_{1x,l} e^{jl\omega}\right|^2}, \tag{19a}$$

$$\gamma_2^{(1)}(\omega) \triangleq \frac{\sigma_2^2 \beta P_1 \left|1 - \sum_{l=1}^{q_2} \kappa_{2,l} e^{jl\omega}\right|^2}{\lambda \left|1 - \sum_{l=1}^{p_2} \rho_{2x,l} e^{jl\omega}\right|^2}, \tag{19b}$$

$$\gamma_1^{(2)}(\omega) \triangleq \frac{\sigma_1^2 \bar{\alpha} P_1 \left|1 - \sum_{l=1}^{q_1} \kappa_{1,l} e^{jl\omega}\right|^2}{\bar{\lambda} \left|1 - \sum_{l=1}^{p_1} \rho_{1x,l} e^{jl\omega}\right|^2}, \tag{19c}$$

$$\gamma_2^{(2)}(\omega) \triangleq \frac{\sigma_2^2 \bar{\beta} P_2 \left|1 - \sum_{l=1}^{q_2} \kappa_{2,l} e^{jl\omega}\right|^2}{\bar{\lambda} \left|1 - \sum_{l=1}^{p_2} \rho_{2x,l} e^{jl\omega}\right|^2}, \tag{19d}$$

where $\sigma_2^2$ is chosen such that $\mathrm{Var}\{x_{2,i}\} = \beta P_2/\lambda$ for all $\lambda 1 \leq i \leq n$.

• One can generalize the above results by using different parameters in the two segments. For example, for the first-order ARMA process, instead of keeping the same parameters $\rho_{x_1}$ and $\kappa_1$ within the whole codeword $\boldsymbol{x}_1$, we can conjugate the parameters $\rho_{1x_1}$ and $\kappa_{11}$ for $i = 1, \ldots, n\lambda$, and a correlation of parameters $\rho_{2x_1}$ and $\kappa_{21}$ for $i = n\lambda + 1, \ldots, n$. Accordingly, one would obtain the same result as before, but replacing every instance of $\gamma_i^{(k)}$ for $i, k = 1, 2$ with:

$$\gamma_1^{(1)}(\omega) \triangleq \frac{\sigma_{11}^2 \alpha P_1 \left|1 - \kappa_{11} e^{j\omega}\right|^2}{\lambda \left|1 - \rho_{1x_1} e^{j\omega}\right|^2}, \tag{20a}$$

$$\gamma_2^{(1)}(\omega) \triangleq \frac{\sigma_{12}^2 \beta P_2 \left|1 - \kappa_{12} e^{j\omega}\right|^2}{\lambda \left|1 - \rho_{1x_2} e^{j\omega}\right|^2}, \tag{20b}$$

$$\gamma_1^{(2)}(\omega) \triangleq \frac{\sigma_{21}^2 \bar{\alpha} P_1 \left|1 - \kappa_{21} e^{j\omega}\right|^2}{\bar{\lambda} \left|1 - \rho_{2x_1} e^{j\omega}\right|^2}, \tag{20c}$$

$$\gamma_2^{(2)}(\omega) \triangleq \frac{\sigma_{22}^2 \bar{\beta} P_2 \left|1 - \kappa_{22} e^{j\omega}\right|^2}{\bar{\lambda} \left|1 - \rho_{2x_2} e^{j\omega}\right|^2}, \tag{20d}$$

where

$$\sigma_{kl}^2 \triangleq \frac{1 - \rho_{kx_l}^2}{1 + \kappa_{kl}^2 - 2\kappa_{kl}\rho_{kx_l}}, \tag{21}$$

and we take the union over all possible allocations of $\{\rho_{kx_l}, \kappa_{kl}\}$ for $k, l \in \{1, 2\}$. Numerical calculations demonstrate that in some cases, this generalization improves the achievable region.

• The capacity region of a discrete memoryless IC was expressed in [12] by the following limiting expression:

$$C_{\text{IC}} = \limsup_{n\to\infty} \text{conv} \bigcup_{P_{X_1}^n, P_{X_2}^n} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq \frac{1}{n} I(X_1^n; Y_1^n) \\ R_2 \leq \frac{1}{n} I(X_2^n; Y_2^n) \end{array} \right\} \tag{22}$$

where limit superior of a sequence of sets, $\{A_n\}$, is defined as $\limsup_{n\to\infty} A_n = \cap_{N=1}^{\infty} \cup_{n\geq N} A_n$. Accordingly, it is tempting to compare the achievable region resulting from (22), when restricting the input distribution to Gaussian ARMA processes. When doing so, numerical examples show that our proposed achievable region is better.

• The various integrals in (17) can actually be calculated by using Cauchy's integral formula, which implies that $\int_0^{2\pi} \log_2\left(1+\zeta e^{j\omega}\right) \mathrm{d}\omega = 0$, whenever $|\zeta| < 1$. For example, for $\kappa_1 = \kappa_2 = 0$, we get:

$$\varphi\left[\gamma_1^{(1)}\right] = \frac{1}{4\pi}\int_0^{2\pi} \log_2\left(1+\gamma_1^{(1)}(\omega)\right)\mathrm{d}\omega \tag{23}$$

$$= \frac{1}{4\pi}\int_0^{2\pi} \log_2\left(\frac{K\cdot\left|1-\zeta e^{j\omega}\right|^2}{|1-\rho_{x_1}e^{j\omega}|^2}\right)\mathrm{d}\omega \tag{24}$$

$$= \frac{1}{2}\log_2 K \tag{25}$$

where

$$K \triangleq \frac{\rho_{x_1}}{\zeta}, \tag{26a}$$

$$\zeta \triangleq \frac{\chi}{2\rho_{x_1}} - \sqrt{\frac{\chi^2}{4\rho_{x_1}^2} - 1}, \tag{26b}$$

$$\chi \triangleq 1 + \rho_{x_1}^2 + \frac{(1-\rho_{x_1}^2)\alpha P_1}{\lambda}. \tag{26c}$$

Unfortunately, the formulas of the other integrals are more complicated, and thus we keep the above integral representations.

In the following, we provide a numerical example which illustrates the improvement compared to [3]. First, note that by construction, $\mathcal{R}_{\text{sason}} \subseteq \mathcal{R}^{(1)}$. Fig. 1 presents a typical comparison between: $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, $\mathcal{R}_{\text{AR}}^{(1)}$ which is the region $\mathcal{R}^{(1)}$ when fixing $\kappa_1 = \kappa_2 = 0$ (an AR process), and $\mathcal{R}_{\text{MA}}^{(1)}$ which is the region when fixing $\rho_{x_1} = \rho_{x_2} = 0$ (an MA process), for $P_1 = 6, P_2 = 1, a_{12} = 3$, and $a_{21} = 0.1$. The numerical calculations were carried out by an exhaustive search over a grid on the parameter space, with a step-size of $10^{-2}$. It is evident from the figure that the new proposed regions (AR, MA, and ARMA) strictly include $\mathcal{R}_{\text{sason}}$. Also, it can be seen that there is a noticeable improvement when using ARMA input processes compared to AR input processes. The improvement compared to MA input processes is less significant. It is interesting, however, to check what are the ARMA parameters which achieve rates that cannot be achieved by the other filters. Consider, for example, the pair $(R_1, R_2) = (1.191, 0.3909)$, which is marked by a small circle in Fig. 1. To achieve this point, one should choose: $\kappa_1 = 0.2605$, $\kappa_2 = 0.9801$, $\rho_{x_1} = 0.7425$, and $\rho_{x_2} = 0.4950$. The moduli of the frequency responses of the filters of the two users, i.e.,

$$H_i(\omega) \triangleq \frac{1-\kappa_i e^{-j\omega}}{1-\rho_{x_i}e^{-j\omega}},\ i = 1, 2, \tag{27}$$

are presented in Fig. 2. It can be seen that the frequency responses for the two users tend to amplify/attenuate in different frequency regions (one of them is a low-pass filter, and the other is a

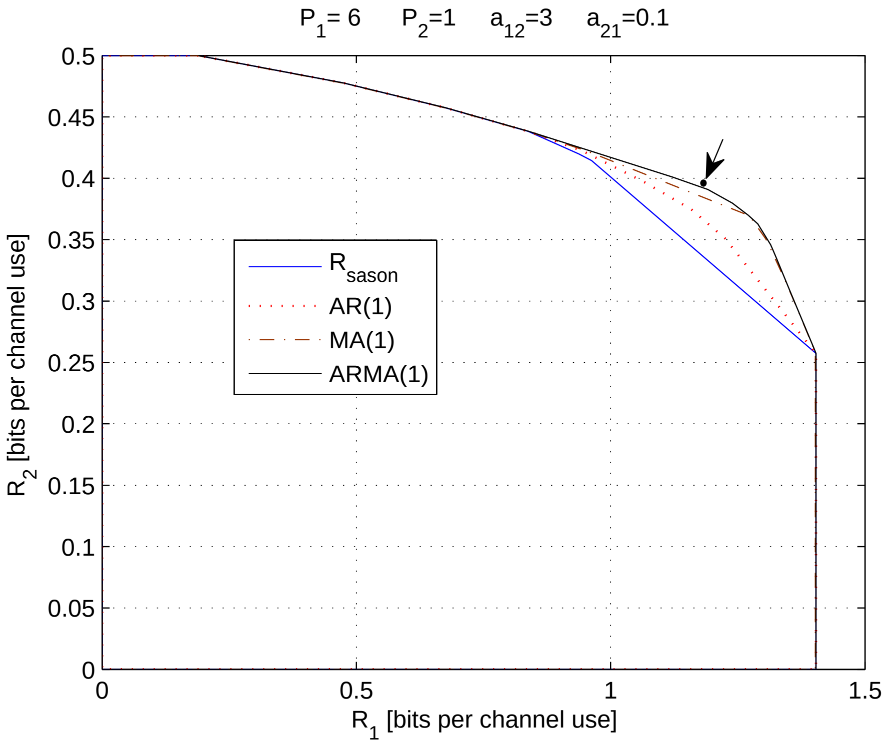


Fig. 1: Comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$ (ARMA), $\mathcal{R}^{(1)}$ when fixing $\kappa_1 = \kappa_2 = 0$ (AR), and $\mathcal{R}^{(1)}$ when fixing $\rho_{x_1} = \rho_{x_2} = 0$ (MA), for $P_1 = 6, P_2 = 1, a_{12} = 3$, and $a_{21} = 0.1$.

high-pass filter), which makes sense.

Fig. 3 also illustrates the improvement compared to [3]. It presents a comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, and $\mathcal{R}^{(1)}_{\text{MA}}$, for $P_1 = 6, P_2 = 1, a_{12} = 2$, and $a_{21} = 0$, that is, a ZIC. It is evident that our region strictly includes that of [3]. Also, it can be seen that our upper corner point is greater than that of $\mathcal{R}_{\text{sason}}$. Note, however, that one can show that the upper corner point of the capacity region, for the given setting, is given by [8]:

$$C_1' \triangleq \max\left\{R_1 : \left(R_1, \frac{1}{2}\log_2\left[1 + P_2\right]\right) \in C_{\text{IC}}\right\} = 1.083, \tag{28}$$

which is larger than our values. This observation motivates us to consider more sophisticated encoding/decoding schemes, described in the following subsection.

## B. *HK-based inner bound:*

In this subsection, we consider encoding/decoding schemes based on the Han-Kobayashi (HK) scheme [4], which will be briefly reviewed in the sequel. The idea is to split the message $M_1$ into "private"

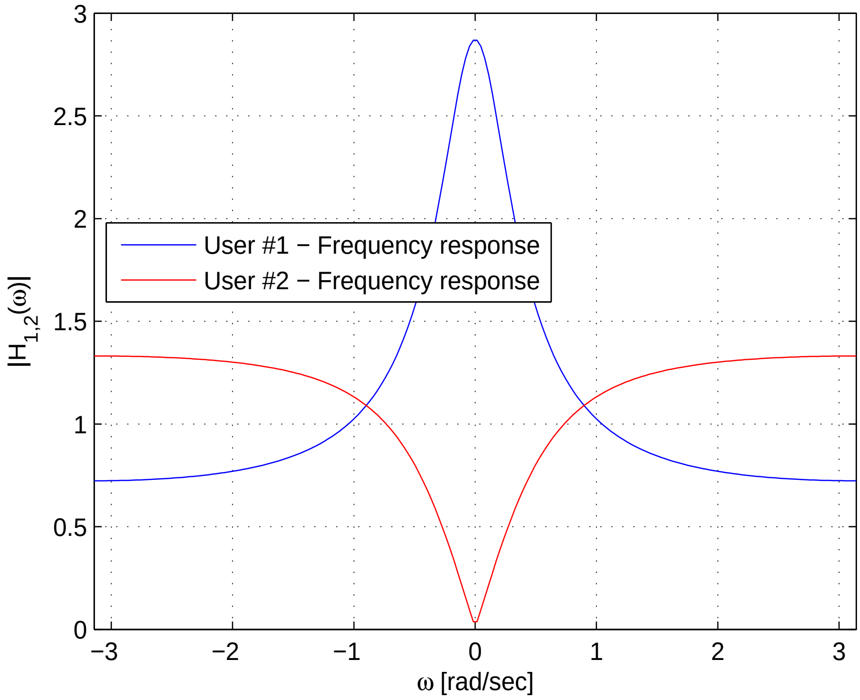


Fig. 2: The modulus of the frequency response of the two users.

and "common" messages, $M_{11}$ and $M_{12}$ at rates $R_{11}$ and $R_{12}$, respectively, such that $R_1 = R_{11} + R_{12}$. Similarly, $M_2$ is split into two messages $M_{21}$ and $M_{22}$, at rates $R_{21}$ and $R_{22}$, respectively, such that $R_2 = R_{21} + R_{22}$. Then, receiver $k = 1, 2$ recovers its intended message $M_k$ and the common message from the other sender (although it is not required to). This scheme is illustrated in Fig. 4. The intuition behind this splitting is based on the receiver's behavior at low and high signal-to-noise ratios (SNRs). Specifically, it is well-known [1] that: (1) when the SNR is low, treating the interference as noise is an optimal strategy, and (2) when the SNR is high, decoding and then canceling the interference is the optimal strategy. Accordingly, the above splitting captures the general intermediate situation, where the first decoder, for example, is interested only in partial information from the second user, in addition to its own intended message. When i.i.d. random coding is employed, it was shown in [4] that the following rate region is achievable:

$$R_1 \leq \rho_1, \;\; R_2 \leq \rho_2, \;\; R_1 + R_2 \leq \rho_{12}, \;\; 2R_1 + R_2 \leq \rho_{10}, \;\; R_1 + 2R_2 \leq \rho_{20}, \tag{29}$$

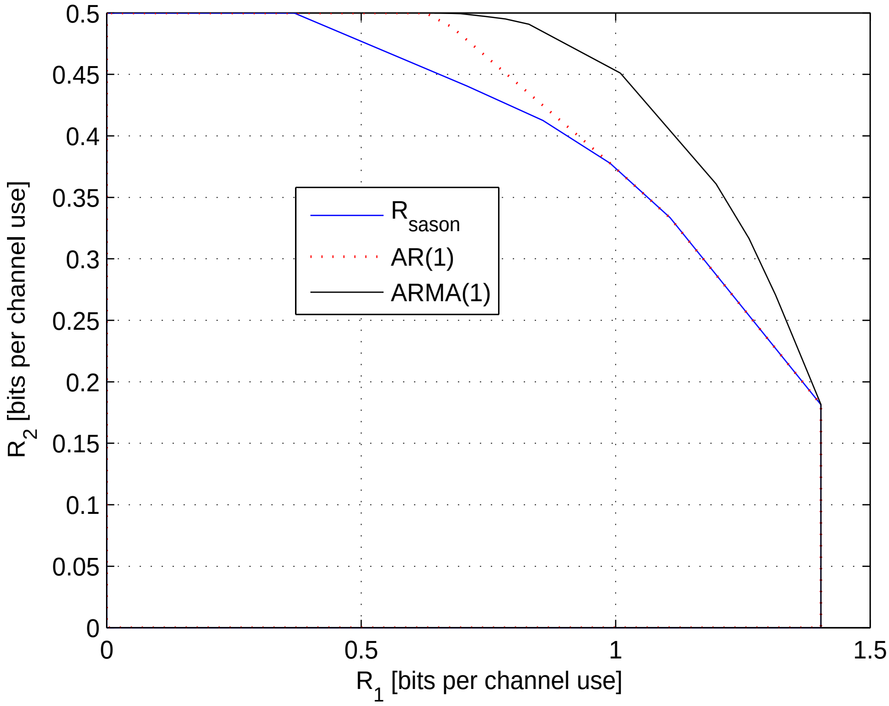


Fig. 3: Comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$ (ARMA), and $\mathcal{R}^{(1)}$ when fixing $\kappa_1 = \kappa_2 = 0$ (AR), for $P_1 = 6, P_2 = 1, a_{12} = 2$, and $a_{21} = 0$.

where:

$$\rho_1 = \sigma_1^* + I(Y_1; U_1|V_1 V_2 Q) \tag{30}$$

$$\rho_2 = \sigma_2^* + I(Y_2; U_2|V_1 V_2 Q) \tag{31}$$

$$\rho_{12} = \sigma_{12}^* + I(Y_1; U_1|V_1 V_2 Q) + I(Y_2; U_2|V_1 V_2 Q) \tag{32}$$

$$\begin{aligned}\rho_{10} = {} & 2\sigma_1^* + 2I(Y_1; U_1|V_1 V_2 Q) + I(Y_2; U_2|V_1 V_2 Q) - [\sigma_1^* - I(Y_2; V_1|V_2 Q)]^+ \\ & + \min\left\{I(Y_2; V_2|V_1, Q), I(Y_2; V_2|Q) + [I(Y_2; V_1|V_2 Q) - \sigma_1^*]^+ ,\right. \\ & \left. +I(Y_1; V_2|V_1, Q), I(Y_1; V_1, V_2|Q) - \sigma_1^*\right\}\end{aligned} \tag{33}$$

$$\begin{aligned}\rho_{20} = {} & 2\sigma_2^* + I(Y_1; U_1|V_1 V_2 Q) + 2I(Y_2; U_2|V_1 V_2 Q) - [\sigma_2^* - I(Y_1; V_2|V_1 Q)]^+ \\ & + \min\left\{I(Y_1; V_1|V_2, Q), I(Y_1; V_1|Q) + [I(Y_1; V_2|V_1 Q) - \sigma_2^*]^+ ,\right. \\ & \left. +I(Y_2; V_1|V_2, Q), I(Y_2; V_2, V_1|Q) - \sigma_2^*\right\},\end{aligned} \tag{34}$$

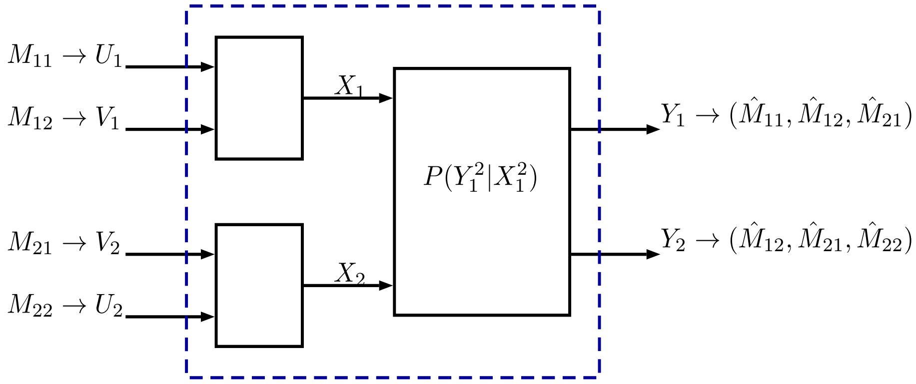


Fig. 4: Han-Kobayashi coding scheme.

and

$$\sigma_1^* = \min\{I(Y_1; V_1|V_2, Q), I(Y_2; V_1|U_2, V_2, Q)\} \tag{35}$$

$$\sigma_2^* = \min\{I(Y_2; V_2|V_1, Q), I(Y_1; V_2|U_1, V_1, Q)\} \tag{36}$$

$$\begin{aligned}\sigma_{12}^* = \min\{&I(Y_1; V_1, V_2|Q), I(Y_2; V_1, V_2|Q), I(Y_1; V_1|V_2, Q) + I(Y_2; V_2|V_1, Q),\\ &I(Y_2; V_1|V_2, Q) + I(Y_1; V_2|V_1, Q)\},\end{aligned} \tag{37}$$

where $U_1, U_2, V_1, V_2$ are auxiliary random variables[3], and $Q$ is a time-sharing variable.

Now, consider the GIC. As was mentioned before, evaluating the HK region for the Gaussian case, is prohibitively complex, due to the auxiliary random variables. Unfortunately, it is still unknown how to choose them optimally. A state-of-the-art common choice in the literature is the following:

$$X_1 = U_1 + V_1, \quad U_1 \sim N(0, \xi_1 P_1),\ V_1 \sim N(0, \bar{\xi}_1 P_1), \tag{38a}$$

$$X_2 = U_2 + V_2, \quad U_2 \sim N(0, \xi_2 P_2),\ V_2 \sim N(0, \bar{\xi}_2 P_2), \tag{38b}$$

where $U_i$ and $V_i$ are statistically independent for $i = 1, 2$, and $\xi_1, \xi_2 \in [0, 1]$ are auxiliary parameters.

In the spirit of the previous subsection, we next provide two achievable regions, which are based on a combination of the Sato and the HK schemes. The difference between these regions will be in their decoding techniques. Also, we emphasize that, as in the previous subsection, here too, one can obtain results for general Gaussian stationary processes. For simplicity of the demonstration, however, in order to keep a reasonable number of parameters to be optimized, we will confine ourselves to first-order

[3]The intuition behind these auxiliaries is that the private messages are transmitted via $U_1$ and $U_2$, while the common messages are transmitted via $V_1$ and $V_2$.

Markov processes (first-order AR). Let us describe briefly the encoding technique.

***Encoding:*** As before, we consider two modes of work. At the first $\lambda$ fraction of the transmission time, the symbols of $\boldsymbol{x}_1$ are as in (13a), and we generate $2^{nR_1}$ such independent codewords $\{\boldsymbol{x}_1(i)\}$. For the second user, however, we use rate-splitting technique, as was used in the HK scheme, and generate $2^{nR_2'}$ and $2^{nR_2''}$ independent codewords $\{\boldsymbol{v}_2(i)\}$ and $\{\boldsymbol{u}_2(i)\}$, respectively, where $0 \leq R_2' \leq R_2$ and $R_2'' = R_2 - R_2'$, in the following way:

$$u_{2,i} = \rho_{u_2} u_{2,i-1} + \sqrt{1-\rho_{u_2}^2} w_{2,i}^u, \tag{39a}$$

$$v_{2,i} = \rho_{v_2} v_{2,i-1} + \sqrt{1-\rho_{v_2}^2} w_{2,i}^v, \tag{39b}$$

where $|\rho_{u_2}|, |\rho_{v_2}| < 1$, $u_{2,0}$ and $v_{2,0}$ are Gaussian random variables with zero mean and variance $\beta\xi_2 P_2/\lambda$ and $\beta\bar{\xi}_2 P_2/\lambda$, respectively, and $\left\{w_{2,i}^u\right\}$ and $\left\{w_{2,i}^v\right\}$ are i.i.d. Gaussian processes with zero mean and variance $\beta\xi_2 P_2/\lambda$, and $\beta\bar{\xi}_2 P_2/\lambda$, respectively, where $0 \leq \beta, \xi_2 \leq 1$. Then, the $2^{nR_2}$ codewords $\{\boldsymbol{x}_2(i,j)\}$ are given by the element-wise addition:

$$\boldsymbol{x}_2(i,j) = \boldsymbol{u}_2(i) + \boldsymbol{v}_2(j) \tag{40}$$

for $i = 1, 2, \ldots, 2^{nR_2''}$ and $j = 1, 2, \ldots, 2^{nR_2'}$. Intuitively, the $\{\boldsymbol{v}_2(i)\}$ codewords serve as the common information to be decoded by both receivers, while $\{\boldsymbol{u}_2(i)\}$ codewords are the private ones, to be decoded only by the second receiver. Finally, during the remaining fraction $\bar{\lambda} = 1 - \lambda$ of the transmission time, the roles of the encoders are swapped. Specifically, the symbols of $\boldsymbol{x}_2$ are as in (15b). For $\boldsymbol{x}_1$, we use rate-splitting, and generate $2^{nR_1'}$ and $2^{nR_1''}$ independent codewords $\{\boldsymbol{v}_1(i)\}$ and $\{\boldsymbol{u}_1(i)\}$, respectively. Then, the $2^{nR_1}$ codewords $\{\boldsymbol{x}_1(i,j)\}$ are given by:

$$\boldsymbol{x}_1(i,j) = \boldsymbol{u}_1(i) + \boldsymbol{v}_1(j) \tag{41}$$

for $i = 1, 2, \ldots, 2^{nR_1''}$ and $j = 1, 2, \ldots, 2^{nR_1'}$.

Now, we consider two decoding techniques:

***Decoding #1 - Successive cancellation decoding:***

- Mode #1: Receiver 1 first decodes the common message $\boldsymbol{v}_2$, and then uses it as side information for decoding $\boldsymbol{x}_1$. Receiver 2, on the other hand, decodes (simultaneously, as with the MAC) his messages $\boldsymbol{v}_2$ and $\boldsymbol{u}_2$. This mode of work will be used a fraction $\lambda$ of the transmission time.
- Mode #2: Receiver 2 first decodes the common message $\boldsymbol{v}_1$, and then uses it as side information for decoding $\boldsymbol{x}_2$. Receiver 1, decodes his messages $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$ simultaneously. This mode of work will be used a fraction $\bar{\lambda}$ of the transmission time.

***Decoding #2 - Simultaneous decoding:***

• Mode #1: Receiver 1 decodes simultaneously the messages $\boldsymbol{v}_2$ and $\boldsymbol{x}_1$. Similarly, receiver 2 decodes simultaneously $\boldsymbol{v}_2$ and $\boldsymbol{u}_2$. This mode of work will be used a fraction $\lambda$ of the transmission time.

• Mode #2: Receiver 2 decodes simultaneously the messages $\boldsymbol{v}_1$ and $\boldsymbol{x}_2$. Similarly, receiver 1 decodes simultaneously $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$. This mode of work will be used a fraction $\bar{\lambda}$ of the transmission time.

The following theorems, which are proved in Appendix C, give the achievable rate regions resulting from the above encoding/decoding schemes. Let us first define the following functions:

$$\gamma_{x_1}(\omega) \triangleq \frac{(1-\rho_{x_1}^2)\alpha P_1}{\lambda\,|1-\rho_{x_1}e^{j\omega}|^2}, \tag{42a}$$

$$\gamma_{x_2}(\omega) \triangleq \frac{(1-\rho_{x_2}^2)\bar{\beta} P_2}{\bar{\lambda}\,|1-\rho_{x_2}e^{j\omega}|^2}, \tag{42b}$$

$$\gamma_{u_1}(\omega) \triangleq \frac{(1-\rho_{u_1}^2)\xi_1\bar{\alpha} P_1}{\bar{\lambda}\,|1-\rho_{u_1}e^{j\omega}|^2}, \tag{42c}$$

$$\gamma_{v_1}(\omega) \triangleq \frac{(1-\rho_{v_1}^2)\bar{\xi}_1\bar{\alpha} P_1}{\bar{\lambda}\,|1-\rho_{v_1}e^{j\omega}|^2}, \tag{42d}$$

$$\gamma_{u_2}(\omega) \triangleq \frac{(1-\rho_{u_2}^2)\xi_2\beta P_2}{\lambda\,|1-\rho_{u_2}e^{j\omega}|^2}, \tag{42e}$$

$$\gamma_{v_2}(\omega) \triangleq \frac{(1-\rho_{v_2}^2)\bar{\xi}_2\beta P_2}{\lambda\,|1-\rho_{v_2}e^{j\omega}|^2}, \tag{42f}$$

where $\omega \in [0,2\pi)$, $|\rho_{u_1}|, |\rho_{v_1}|, |\rho_{u_2}|, |\rho_{v_2}| < 1$, and $\xi_1, \xi_2 \in [0,1]$. Finally, define $\boldsymbol{\xi} = (\alpha, \beta, \lambda, \xi_1, \xi_2, \rho_{x_1}, \rho_{x_2}, \rho_{u_1}, \rho_{v_1}, \rho_{u_2}, \rho_{v_2})$, and $\boldsymbol{\Xi} \triangleq [0,1]^5 \times (-1,1)^6$.

*Theorem 4 (Decoding #1)* The set of rate pairs:

$$\mathcal{R}^{(2)} = \bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \left\{ (R_1,R_2) : \begin{matrix} R_1 \le R_1(\boldsymbol{\xi}) \\ R_2 \le R_2(\boldsymbol{\xi}) \end{matrix} \right\} \tag{43}$$

where

$$\begin{aligned} R_1(\boldsymbol{\xi}) \triangleq \lambda\cdot\varphi\left[\frac{\gamma_{x_1}}{1+a_{12}\gamma_{u_2}}\right] + \bar{\lambda}\min\Bigg\{ &\varphi\left[\frac{\gamma_{u_1}+\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right], \varphi\left[\frac{\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right] + \varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right], \\ &\varphi\left[\frac{a_{21}\gamma_{v_1}}{1+\gamma_{x_2}+a_{21}\gamma_{u_1}}\right] + \varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right]\Bigg\}, \end{aligned} \tag{44}$$

and

$$\begin{aligned} R_2(\boldsymbol{\xi}) \triangleq \bar{\lambda}\cdot\varphi\left[\frac{\gamma_{x_2}}{1+a_{21}\gamma_{u_1}}\right] + \lambda\cdot\min\Bigg\{ &\varphi\left[\frac{\gamma_{u_2}+\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right], \varphi\left[\frac{\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right] + \varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right], \\ &\varphi\left[\frac{a_{12}\gamma_{v_2}}{1+\gamma_{x_1}+a_{12}\gamma_{u_2}}\right] + \varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right]\Bigg\}. \end{aligned} \tag{45}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

*Theorem 5 (Decoding #2)* The set of rate pairs:

$$\mathcal{R}^{(3)} = \bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \left\{ (R_1, R_2) : \begin{array}{c} R_1 \leq \bar{R}_1(\boldsymbol{\xi}) \\ R_2 \leq \bar{R}_2(\boldsymbol{\xi}) \\ R_1 + R_2 \leq R_{\text{sum}}(\boldsymbol{\xi}) \end{array} \right\} \tag{46}$$

where

$$\begin{aligned} \bar{R}_1(\boldsymbol{\xi}) \triangleq \lambda \cdot \varphi\left[\frac{\gamma_{x_1}}{1+a_{12}\gamma_{u_2}}\right] + \bar{\lambda}\min\Bigg\{ & \varphi\left[\frac{\gamma_{u_1}+\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right], \varphi\left[\frac{\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right] + \varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right], \\ & \varphi\left[\frac{a_{21}\gamma_{v_1}}{1+a_{21}\gamma_{u_1}}\right] + \varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right]\Bigg\}, \end{aligned} \tag{47}$$

and

$$\begin{aligned} \bar{R}_2(\boldsymbol{\xi}) \triangleq \bar{\lambda} \cdot \varphi\left[\frac{\gamma_{x_2}}{1+a_{21}\gamma_{u_1}}\right] + \bar{\lambda}\min\Bigg\{ & \varphi\left[\frac{\gamma_{u_2}+\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right], \varphi\left[\frac{\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right] + \varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right], \\ & \varphi\left[\frac{a_{12}\gamma_{v_2}}{1+a_{12}\gamma_{u_2}}\right] + \varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right]\Bigg\}, \end{aligned} \tag{48}$$

and

$$\begin{aligned} R_{\text{sum}}(\boldsymbol{\xi}) \triangleq \lambda \cdot & \left\{\varphi\left[\frac{\gamma_{x_1}+a_{12}\gamma_{v_2}}{1+a_{12}\gamma_{u_2}}\right] + \varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right]\right\} \\ & + \bar{\lambda} \cdot \left\{\varphi\left[\frac{\gamma_{x_2}+a_{21}\gamma_{v_1}}{1+a_{21}\gamma_{u_1}}\right] + \varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right]\right\}, \end{aligned} \tag{49}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

Finally, in view of Theorems 4 and 5, we obtain the following immediate result.

*Corollary 1* The rate region:

$$\mathcal{R}^{(4)} = \text{conv}\left\{\mathcal{R}^{(2)} \cup \mathcal{R}^{(3)}\right\}, \tag{50}$$

is achievable for the two-user GIC in (1) under the power constraints (2).

One may realize that the above encoding scheme can be easily generalized by rate splitting both users in both segments, and not just one user in each segment, as we did above. This modification adds, of course, more parameters to be optimized. Nonetheless, numerical calculations show that there is no noticeable improvement due to this generalization.

Given the above results, let us consider two numerical examples. First, we consider the same example as before, and see the improvement of the new achievable regions. Fig. 5 presents a comparison between

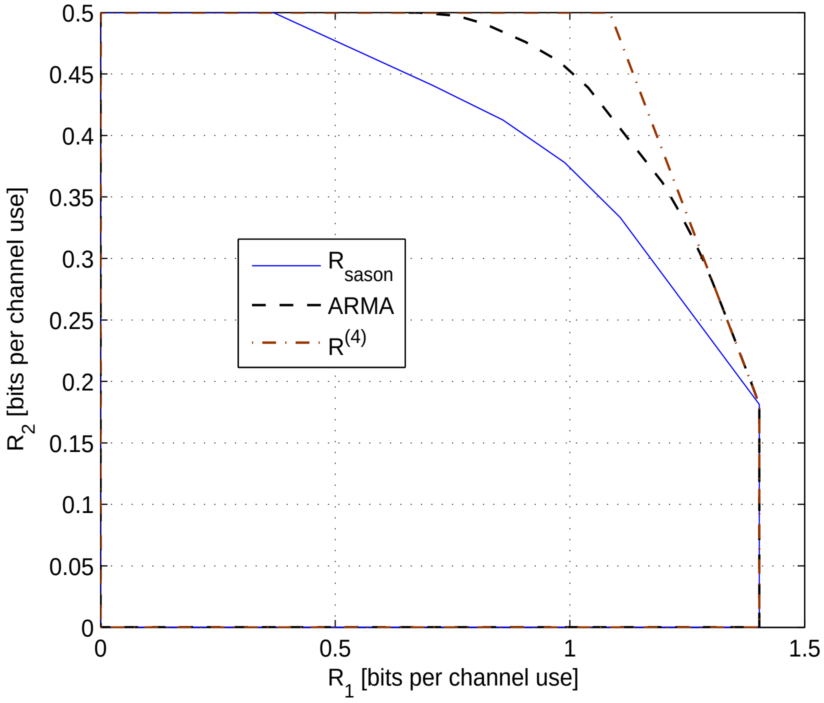


Fig. 5: Comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, and $\mathcal{R}^{(4)}$, for $P_1 = 6, P_2 = 1, a_{12} = 2$, and $a_{21} = 0$.

$\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, and $\mathcal{R}^{(4)}$, for $P_1 = 6, P_2 = 1, a_{12} = 2$, and $a_{21} = 0$. The improvement resulting from the new achievable regions is evident, where $\mathcal{R}^{(3)}$ achieves the best results. Also, note that $\mathcal{R}^{(3)}$ achieves the upper corner point of the capacity region, which is given in (28). Fig. 6 presents a comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, $\mathcal{R}^{(4)}$, and $\mathcal{R}_{\text{HK}}$ for $P_1 = 1, P_2 = 6, a_{12} = 0.1$, and $a_{21} = 0.5$. The region $\mathcal{R}_{\text{HK}}$ is the (state-of-the-art) HK region given in (29) and (38). Here, $\mathcal{R}_{\text{sason}}$ and $\mathcal{R}^{(1)}$ coincide. It can be seen that for this example $\mathcal{R}_{\text{sason}}$ is in some regions better than $\mathcal{R}_{\text{HK}}$, but not everywhere. The new region $\mathcal{R}^{(4)}$, however, uniformly outperforms $\mathcal{R}_{\text{sason}}$ and $\mathcal{R}_{\text{HK}}$.

## III. Conclusion

In this paper, we analyzed several encoding/decoding schemes for the two-user GIC. Usually, as in other multi-user communication systems, achievable rate regions for the GIC are based on ensembles of i.i.d. codewords. In this work, however, we analyzed the impact of using random coding ensembles of codewords with memory, and we show that it can noticeably improve known results which are based on ensembles of i.i.d. codewords. Specifically, we took known achievable rate regions for the GIC, and generalized them by allowing dependency between the code symbols. Numerical calculations show that

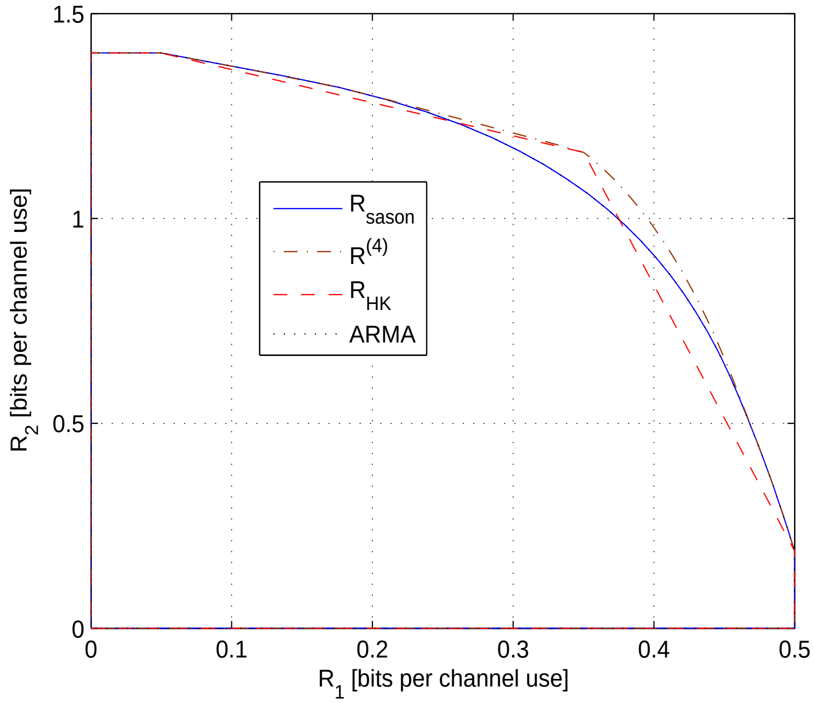


Fig. 6: Comparison between $\mathcal{R}_{\text{sason}}$, $\mathcal{R}^{(1)}$, $\mathcal{R}^{(4)}$, and $\mathcal{R}_{\text{HK}}$ for $P_1 = 1, P_2 = 6, a_{12} = 0.1$, and $a_{21} = 0.5$.

even for very simple memory structures, such as first-order MA process, the obtained achievable rate regions are wider than other known achievable rate regions, and in particular, the state-of-the-art HK region.

The main difficulty with our approach, is the optimization required to obtain the achievable region. Indeed, recall that the optimization is over the filter coefficients (and the time sharing parameters), which may be as large as we wish, but at the expense of computational complexity. A possible simplification is to consider the same filters but with a large number of random coefficients (in the spirit of random coding), distributed according to some given prior distribution. Using this approach, we will end up with optimizing over only three parameters $(\alpha, \beta, \lambda)$, and we can use a large number of filter coefficients. Unfortunately, numerical calculations show that the above randomization approach significantly degrades the achievable regions resulting from the optimizations. Another possible approach is to use time-variant random filters. For example, $\boldsymbol{x}_1 = \boldsymbol{H}_1 \boldsymbol{u}_1$ where $\boldsymbol{H}_1$ is an i.i.d. random matrix, and $\boldsymbol{u}_1$ is an i.i.d. Gaussian random vector, and we choose the variance of $\boldsymbol{u}_1$ to apply the power constraint within each segment. We do the same for $\boldsymbol{x}_2$. Using random matrix theory, an achievable rate region can be derived.

Unfortunately, again, numerical calculations show that this approach fails compared to the optimization over a relatively small number of parameters, as was carried out in the body of this work.

## APPENDIX A
## PROOF OF THEOREM 2

To prove Theorem 2, we will first generalize Sato's result in (6)-(8), such that it will apply to any input distributions, and not just i.i.d.. To this end, we need to analyze the probability of error. We next show that a general achievable rate region, is given by:

$$\tilde{\mathcal{G}}_B = \text{conv}\left\{\tilde{\mathcal{G}}_{B_1} \cup \tilde{\mathcal{G}}_{B_2}\right\} \tag{A.1}$$

where

$$\tilde{\mathcal{G}}_{B_1} \triangleq \lim_{n\to\infty} \text{conv} \bigcup_{P^n_{X_1}, P^n_{X_2}} \left\{ (R_1, R_2) : \begin{array}{c} 0 \le R_1 \le \frac{1}{n} I(X_1^n; Y_1^n | X_2^n) \\ 0 \le R_2 \le \min\left\{\frac{1}{n} I(X_2^n; Y_1^n), \frac{1}{n} I(X_2^n; Y_2^n)\right\} \end{array} \right\} \tag{A.2}$$

and

$$\tilde{\mathcal{G}}_{B_2} \triangleq \lim_{n\to\infty} \text{conv} \bigcup_{P^n_{X_1}, P^n_{X_2}} \left\{ (R_1, R_2) : \begin{array}{c} 0 \le R_1 \le \min\left\{\frac{1}{n} I(X_1^n; Y_1^n), \frac{1}{n} I(X_1^n; Y_2^n)\right\} \\ 0 \le R_2 \le \frac{1}{n} I(X_2^n; Y_2^n | X_1^n) \end{array} \right\}. \tag{A.3}$$

Let $P_1^n(\boldsymbol{x}_1)$ and $P_2^n(\boldsymbol{x}_2)$ be arbitrary probability assignments on the two sets of channel input sequences $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of length $n$. Select independently $M_1$ and $M_2$ codewords $\boldsymbol{x}_{1i}$, for $i = 1, 2, \ldots, M_1$, and $\boldsymbol{x}_{2j}$, for $i = 1, 2, \ldots, M_2$, according to $P_1^n(\boldsymbol{x}_1)$ and $P_2^n(\boldsymbol{x}_2)$, respectively. We now describe the operation of the two decoders corresponding to the region $\tilde{\mathcal{G}}_{B_1}$. Assume that both decoders know the probability distributions $P_1^n(\boldsymbol{x}_1)$ and $P_2^n(\boldsymbol{x}_2)$, and therefore know the following conditional probabilities,

$$P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_1, \boldsymbol{x}_2) P_1^n(\boldsymbol{x}_1) \tag{A.4}$$

$$P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_1, \boldsymbol{x}_2) P_1^n(\boldsymbol{x}_1). \tag{A.5}$$

Decoder 2 chooses the message that has the largest $P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)$ among the $M_2$ codewords. Decoder 1, first chooses the message that has the largest $P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_2)$ among the $M_2$ codewords and then, by using $\boldsymbol{y}_1$ and the decoded $\boldsymbol{x}_2$, chooses the message that has the largest $P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_1, \boldsymbol{x}_2)$ among the $M_1$ codewords. Note that these decoding rules are not the optimal decoding rules (i.e., maximum-likelihood), because they use the conditional probabilities averaged over the random coding distributions. Let

$$P_{e,ij} \triangleq \mathbb{P}\left\{\text{error}|i, j, \boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{y}_1, \boldsymbol{y}_2\right\}, \tag{A.6}$$

be the average probability of decoding error conditioned on the messages $i$ and $j$, the codewords $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{2j}$, and on $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. Then the average probability of decoding error for $(i,j)$ message pair is

$$\bar{P}_{e,ij} = \sum_{\boldsymbol{x}_{1i},\boldsymbol{x}_{2j},\boldsymbol{y}_1,\boldsymbol{y}_2} P_1^n(\boldsymbol{x}_{1i})P_2^n(\boldsymbol{x}_{2j})P^n(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}_{1i},\boldsymbol{x}_{2j})P_{e,ij}. \tag{A.7}$$

The above conditional error event is the union of the three conditional error events, $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$, where: $\mathcal{E}_1$ refers to an error in decoder 2, $\mathcal{E}_2$ refers to an error in decoder 1 for decoding $j$, and $\mathcal{E}_3$ refers to an error in decoder 1 for decoding $i$, but correctly decoding index $j$. Then we have

$$P_{e,ij} \leq \mathbb{P}\left\{\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3\right\} \tag{A.8}$$

$$\leq \mathbb{P}\left\{\mathcal{E}_1\right\} + \mathbb{P}\left\{\mathcal{E}_2\right\} + \mathbb{P}\left\{\mathcal{E}_2^c \cap \mathcal{E}_3\right\} \tag{A.9}$$

$$\leq \mathbb{P}\left\{\mathcal{E}_1\right\} + \mathbb{P}\left\{\mathcal{E}_2\right\} + \mathbb{P}\left\{\mathcal{E}_3|\mathcal{E}_2^c\right\}. \tag{A.10}$$

Using Gallager's bounding technique, we can obtain upper bounds on each of the terms at the r.h.s. of (A.10), and thus obtain:

$$\bar{P}_{e,ij} \leq A + B + C \tag{A.11}$$

where

$$A = (M_2-1)^\rho \sum_{\boldsymbol{y}_2} \left\{ \sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2) P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)^{1/(1+\rho)} \right\}^{1+\rho} \tag{A.12}$$

$$B = (M_2-1)^\rho \sum_{\boldsymbol{y}_1} \left\{ \sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2) P_2^n(\boldsymbol{y}_1|\boldsymbol{x}_2)^{1/(1+\rho)} \right\}^{1+\rho} \tag{A.13}$$

$$C = (M_1-1)^\rho \sum_{\boldsymbol{y}_1,\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2) \left\{ \sum_{\boldsymbol{x}_1} P_1^n(\boldsymbol{x}_1) P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_1\boldsymbol{x}_2)^{1/(1+\rho)} \right\}^{1+\rho}. \tag{A.14}$$

Indeed, let us derive for example the bound on $\mathbb{P}\left\{\mathcal{E}_1\right\}$. We have:

$$\mathbb{P}\left\{\mathcal{E}_1\right\} = \frac{1}{M_2} \sum_i \sum_{\boldsymbol{x}_{2,i},\boldsymbol{y}_2} P_2^n(\boldsymbol{x}_{2,i}) P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i}) \mathbb{P}\left\{\text{error}|\boldsymbol{x}_{2,i},\boldsymbol{y}_2\right\}. \tag{A.15}$$

Now, define the event $\mathcal{A}_{i'}$ for each $i' \neq i$, as the event that codeword $\boldsymbol{x}_{2,i'}$ is selected, that is, $P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i'}) \geq P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})$. Thus, we have:

$$\mathbb{P}\left\{\text{error}|\boldsymbol{x}_{2,i},\boldsymbol{y}_2\right\} \leq \mathbb{P}\left\{ \bigcup_{i'\neq i} \mathcal{A}_{i'} \right\} \leq \left[ \sum_{i'\neq i} \mathbb{P}\left\{\mathcal{A}_{i'}\right\} \right]^\rho \tag{A.16}$$

for any $0 < \rho \leq 1$. From the definition of $\mathcal{A}_{i'}$, we have:

$$\mathbb{P}\left\{\mathcal{A}_{i'}\right\} = \sum_{\mathcal{A}_{i'}} P_2^n(\boldsymbol{x}_{2,i'}) \tag{A.17}$$

$$\leq \sum_{\boldsymbol{x}_{2,i'}} P_2^n(\boldsymbol{x}_{2,i'}) \left[\frac{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i'})}{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})}\right]^s \tag{A.18}$$

for $s > 0$. Since $\boldsymbol{x}_{2,i'}$ is a dummy variable of the summation in (A.18), the subscript $i'$ can be dropped and the bound is independent of $i'$. Hence,

$$\mathbb{P}\left\{\text{error}|\boldsymbol{x}_{2,i},\boldsymbol{y}_2\right\} \leq \left[(M_2-1)\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)\left[\frac{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)}{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})}\right]^s\right]^\rho. \tag{A.19}$$

Therefore, substituting the above in (A.15), we obtain:

$$\mathbb{P}\left\{\mathcal{E}_1\right\} \leq (M_2-1)^\rho \frac{1}{M_2}\sum_i \sum_{\boldsymbol{x}_{2,i},\boldsymbol{y}_2} P_2^n(\boldsymbol{x}_{2,i})P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})\left[\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)\left[\frac{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)}{P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})}\right]^s\right]^\rho \tag{A.20}$$

$$= (M_2-1)^\rho \frac{1}{M_2}\sum_i \sum_{\boldsymbol{y}_2}\left[\sum_{\boldsymbol{x}_{2,i}} P_2^n(\boldsymbol{x}_{2,i})P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_{2,i})^{1-s\rho}\right]\left[\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)^s\right]^\rho. \tag{A.21}$$

Finally, we substitute $s = 1/(1+\rho)$, and we note that $\boldsymbol{x}_{2,i}$ is a dummy variable of summation, so we obtain:

$$\mathbb{P}\left\{\mathcal{E}_1\right\} \leq (M_2-1)^\rho \sum_{\boldsymbol{y}_2}\left[\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)^{1/(1+\rho)}\right]^{1+\rho}, \tag{A.22}$$

which is (A.12). In the same way, we can obtain (A.13) and (A.14).

Returning to (A.10), by maximizing over $\rho$ we obtain:

$$A = 2^{-nE_A(R_2)} \quad B = 2^{-nE_B(R_2)} \quad C = 2^{-nE_C(R_1)}, \tag{A.23}$$

where $R_1 = (\log_2 M_1)/n$, $R_2 = (\log_2 M_2)/n$, and

$$E_A(R_2) = \max_{0\leq\rho\leq 1}\left[E_{A_0}(\rho) - \rho R_2\right] \tag{A.24}$$

$$E_B(R_2) = \max_{0\leq\rho\leq 1}\left[E_{B_0}(\rho) - \rho R_2\right] \tag{A.25}$$

$$E_C(R_1) = \max_{0\leq\rho\leq 1}\left[E_{C_0}(\rho) - \rho R_1\right] \tag{A.26}$$

in which

$$E_{A_0}(\rho) = -\frac{1}{n}\log_2 \sum_{\boldsymbol{y}_2}\left\{\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)P_2^n(\boldsymbol{y}_2|\boldsymbol{x}_2)^{1/(1+\rho)}\right\}^{1+\rho} \tag{A.27}$$

$$E_{B_0}(\rho) = -\frac{1}{n}\log \sum_{\boldsymbol{y}_1}\left\{\sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)P_2^n(\boldsymbol{y}_1|\boldsymbol{x}_2)^{1/(1+\rho)}\right\}^{1+\rho} \tag{A.28}$$

$$E_{C_0}(\rho) = -\frac{1}{n}\log \sum_{\boldsymbol{y}_1,\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2)\left\{\sum_{\boldsymbol{x}_1} P_1^n(\boldsymbol{x}_1)P_1^n(\boldsymbol{y}_1|\boldsymbol{x}_1,\boldsymbol{x}_2)^{1/(1+\rho)}\right\}^{1+\rho}. \tag{A.29}$$

Since the average probability of error $\bar{P}_{e,ij}$ does not depend on $i$ and $j$, we may write:

$$\bar{P}_e \leq 2^{-nE_A(R_2)} + 2^{-nE_B(R_2)} + 2^{-nE_C(R_1)}. \tag{A.30}$$

Now, as usual, one can show in the usual way that the error exponents are positive if:

$$R_1 < \frac{1}{n} I(X_1^n; Y_1^n | X_2^n), \tag{A.31}$$

$$R_2 < \min\left\{\frac{1}{n} I(X_2^n; Y_1^n), \frac{1}{n} I(X_2^n; Y_2^n)\right\}, \tag{A.32}$$

where

$$I(X_1^n; Y_1^n | X_2^n) = \sum_{\boldsymbol{x}_2} P_2^n(\boldsymbol{x}_2) \sum_{\boldsymbol{x}_1, \boldsymbol{y}_1} P_1^n(\boldsymbol{x}_1) P^n(\boldsymbol{y}_1 | \boldsymbol{x}_1, \boldsymbol{x}_2) \log \frac{P^n(\boldsymbol{y}_1 | \boldsymbol{x}_1, \boldsymbol{x}_2)}{P_1^n(\boldsymbol{y}_1 | \boldsymbol{x}_2)}, \tag{A.33}$$

$$I(X_2^n; Y_1^n) = \sum_{\boldsymbol{x}_2, \boldsymbol{y}_1} P_2^n(\boldsymbol{x}_2) P_1^n(\boldsymbol{y}_1 | \boldsymbol{x}_2) \log \frac{P_1^n(\boldsymbol{y}_1 | \boldsymbol{x}_2)}{P^n(\boldsymbol{y}_1)}, \tag{A.34}$$

$$I(X_2^n; Y_2^n) = \sum_{\boldsymbol{x}_2, \boldsymbol{y}_2} P_2^n(\boldsymbol{x}_2) P_2^n(\boldsymbol{y}_2 | \boldsymbol{x}_2) \log \frac{P_2^n(\boldsymbol{y}_2 | \boldsymbol{x}_2)}{P^n(\boldsymbol{y}_2)}, \tag{A.35}$$

where, again, $P_1^n(\boldsymbol{x}_1)$ and $P_2^n(\boldsymbol{x}_2)$ are the random coding distributions. Note that the derivatives of $E_{A_0}(\rho)$, $E_{B_0}(\rho)$, and $E_{C_0}(\rho)$, w.r.t. $\rho$, evaluated at $\rho = 0$, give the above mutual information terms. Therefore, we have proved that the rate-pair within the region $\tilde{\mathcal{G}}_{B_1}$ is achievable. The proof for $\tilde{\mathcal{G}}_{B_2}$ is similar. Combining these two regions by time-sharing, we complete the proof.

Given the above result, we are ready to prove Theorem 2. Consider the situation where during a fraction $\lambda$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ form stationary Gaussian processes with zero mean, and Toeplitz covariance matrices $\tilde{\boldsymbol{R}}_{x_1}^{(1)} \triangleq (\alpha P_1/\lambda) \cdot \boldsymbol{R}_{x_1}^{(1)}$ and $\tilde{\boldsymbol{R}}_{x_2}^{(1)} \triangleq (\beta P_2/\lambda) \cdot \boldsymbol{R}_{x_2}^{(1)}$, with entries $\{r_{11,i}\}_{i=0}^{n-1}$ and $\{r_{21,i}\}_{i=0}^{n-1}$, respectively, where $r_{11,0} = r_{21,0} = 1$, and $|r_{11,i}|, |r_{21,i}| \leq 1$ for $i > 1$. During the remaining fraction $\bar{\lambda} = 1 - \lambda$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ form again stationary Gaussian processes with zero mean, and Toeplitz covariance matrices $\tilde{\boldsymbol{R}}_{x_1}^{(2)} \triangleq (\bar{\alpha} P_1/\bar{\lambda}) \boldsymbol{R}_{x_1}^{(2)}$ and $\tilde{\boldsymbol{R}}_{x_2}^{(2)} \triangleq (\bar{\beta} P_2/\bar{\lambda}) \boldsymbol{R}_{x_2}^{(2)}$ with entries $\{r_{12,i}\}_{i=0}^{n-1}$ and $\{r_{22,i}\}_{0=1}^{n-1}$, respectively, where $r_{12,0} = r_{22,0} = 1$, and $|r_{12,i}|, |r_{22,i}| \leq 1$ for $i > 1$. Finally, note that the two input codewords satisfy the power constraints,

$$\frac{1}{n} \mathbb{E} \|\boldsymbol{x}_1\|^2 = \lambda \frac{\alpha P_1}{\lambda} + \bar{\lambda} \frac{\bar{\alpha} P_1}{\bar{\lambda}} = P_1, \tag{A.36}$$

and similarly for the second user. Next, consider two modes of work:

- Mode #1: receiver 1 first decodes the message of the second sender, and then uses it as side information for decoding his message (i.e., subtracts $\sqrt{a_{12}} \boldsymbol{x}_2$ from the received signal $\boldsymbol{y}_1$). On the other hand, receiver 2 directly decodes $\boldsymbol{x}_2$. This mode of work corresponds to the achievable rate region $\tilde{\mathcal{G}}_{B_1}$ in (A.2), and will be used here a fraction $\lambda$ of the transmission time with inputs that are distributed as described above.

• Mode #2: In the second mode (which is dual to the first mode), we refer to the mode of work which corresponds to the achievable rate region $\tilde{\mathcal{G}}_{B_2}$ in (A.3), and assume that it is used during the remaining fraction $\bar{\lambda}$ of the transmission time. We will assume here that during the second mode, the two inputs are distributed as described above.

Now, given the above modes, let $R_1^{(i)}$ and $R_2^{(i)}$ be the transmission rates in mode $i$. Accordingly, by a time-sharing argument, the transmission rates of the two users are

$$(R_1, R_2) = \lambda \cdot (R_1^{(1)}, R_2^{(1)}) + \bar{\lambda} \cdot (R_1^{(2)}, R_2^{(2)}). \tag{A.37}$$

Let us now calculate $R_1^{(i)}$ and $R_2^{(i)}$, for $i = 1, 2$. For $i = 1$, we wish to calculate the mutual information terms in $\tilde{\mathcal{G}}_{B_1}$, i.e.,

$$R_1^{(1)} = \lim_{n\to\infty} \frac{I(X_1^{n\lambda}; Y_1^{n\lambda}|X_2^{n\lambda})}{n\lambda} \tag{A.38}$$

$$R_2^{(1)} = \min\left\{ \lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_1^{n\lambda})}{n\lambda}, \lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_2^{n\lambda})}{n\lambda} \right\} \tag{A.39}$$

Note that:

$$I(X_1^{n\lambda}; Y_1^{n\lambda}|X_2^{n\lambda}) = H(Y_1^{n\lambda}|X_2^{n\lambda}) - H(Y_1^{n\lambda}|X_1^{n\lambda}, X_2^{n\lambda}) \tag{A.40}$$

$$= H(X_1^{n\lambda} + Z_1^{n\lambda}) - H(Z_1^{n\lambda}) \tag{A.41}$$

$$= H(X_1^{n\lambda} + Z_1^{n\lambda}) - \frac{n\lambda}{2}\log(2\pi e). \tag{A.42}$$

Now, $(X_1^{n\lambda} + Z_1^{n\lambda})$ is a Gaussian random vector with zero mean and covariance matrix $\boldsymbol{I} + \tilde{\boldsymbol{R}}_{x_1}^{(1)}$. Thus,

$$I(X_1^{n\lambda}; Y_1^{n\lambda}|X_2^{n\lambda}) = \frac{1}{2}\log\det\left[2\pi e(\boldsymbol{I} + \tilde{\boldsymbol{R}}_{x_1}^{(1)})\right] - \frac{n\lambda}{2}\log(2\pi e) \tag{A.43}$$

$$= \frac{1}{2}\log\det\left[\boldsymbol{I} + \tilde{\boldsymbol{R}}_{x_1}^{(1)}\right]. \tag{A.44}$$

Similarly,

$$I(X_2^{n\lambda}; Y_1^{n\lambda}) = H(X_1^{n\lambda} + \sqrt{a_{12}}X_2^{n\lambda} + Z_1^{n\lambda}) - H(X_1^{n\lambda} + Z_1^{n\lambda}) \tag{A.45}$$

$$= \frac{1}{2}\log\det\left[\boldsymbol{I} + \tilde{\boldsymbol{R}}_{x_1}^{(1)} + a_{12}\tilde{\boldsymbol{R}}_{x_2}^{(1)}\right] - \frac{1}{2}\log\det\left[\boldsymbol{I} + \tilde{\boldsymbol{R}}_{x_1}^{(1)}\right], \tag{A.46}$$

and

$$I(X_2^{n\lambda}; Y_2^{n\lambda}) = H(\sqrt{a_{21}}X_1^{n\lambda} + X_2^{n\lambda} + Z_1^{n\lambda}) - H(\sqrt{a_{21}}X_1^{n\lambda} + N_2^{n\lambda}) \tag{A.47}$$

$$= \frac{1}{2}\log\det\left[\boldsymbol{I} + a_{21}\tilde{\boldsymbol{R}}_{x_1}^{(1)} + \tilde{\boldsymbol{R}}_{x_2}^{(1)}\right] - \frac{1}{2}\log\det\left[\boldsymbol{I} + a_{21}\tilde{\boldsymbol{R}}_{x_1}^{(1)}\right]. \tag{A.48}$$

Next, we calculate the limits of these terms as $n \to \infty$. To this end, we note that due to the stationarity, the input covariance matrices are Toeplitz matrices, and thus we can apply Szegö's theorem [19-22]. Let

$\left\{\rho_{x_i,k}^{(j)}\right\}$ be the eigenvalues of $\tilde{\boldsymbol{R}}_{x_i}^{(j)}$, for $i,j=1,2$. Recall the definitions in (10). Then, using Szegö's theorem we can show that [19, Th. 4.1]:

$$\lim_{n\to\infty}\sum_{k=0}^{n-1}F(\rho_{x_1,k}^{(1)})=\frac{1}{2\pi}\int_0^{2\pi}F(\nu_{11}(\omega))\mathrm{d}\omega \tag{A.49}$$

for any continuous function $F(\cdot)$. Let us apply this result to our problem. First, note that:

$$I(X_1^{n\lambda};Y_1^{n\lambda}|X_2^{n\lambda})=\frac{1}{2}\log\det\left[\boldsymbol{I}+\tilde{\boldsymbol{R}}_{x_1}^{(1)}\right] \tag{A.50}$$

$$=\frac{1}{2}\sum_{k=0}^{n\lambda-1}\log(1+\rho_{x_1,k}^{(1)}). \tag{A.51}$$

Thus, we have:

$$\lim_{n\to\infty}\frac{I(X_1^{n\lambda};Y_1^{n\lambda}|X_2^{n\lambda})}{n\lambda}=\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\nu_{11}(\omega)\right)\mathrm{d}\omega. \tag{A.52}$$

In a similar manner:

$$I(X_2^{n\lambda};Y_1^{n\lambda})=\frac{1}{2}\log\det\left[\boldsymbol{I}+\tilde{\boldsymbol{R}}_{x_1}^{(1)}+a_{12}\tilde{\boldsymbol{R}}_{x_2}^{(1)}\right]-\frac{1}{2}\log\det\left[\boldsymbol{I}+\tilde{\boldsymbol{R}}_{x_1}^{(1)}\right] \tag{A.53}$$

$$=\frac{1}{2}\sum_{k=0}^{n\lambda-1}\log\left(\frac{1+\rho_{x_1,k}^{(1)}+a_{12}\rho_{x_2,k}^{(1)}}{1+\rho_{x_1,k}^{(1)}}\right) \tag{A.54}$$

$$=\frac{1}{2}\sum_{k=0}^{n\lambda-1}\log\left(1+\frac{a_{12}\rho_{x_2,k}^{(1)}}{1+\rho_{x_1,k}^{(1)}}\right). \tag{A.55}$$

Therefore:

$$\lim_{n\to\infty}\frac{I(X_2^{n\lambda};Y_1^{n\lambda})}{n\lambda}=\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{a_{12}\nu_{21}(\omega)}{1+\nu_{11}(\omega)}\right)\mathrm{d}\omega. \tag{A.56}$$

Similarly:

$$\lim_{n\to\infty}\frac{I(X_2^{n\lambda};Y_2^{n\lambda})}{n\lambda}=\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{\nu_{21}(\omega)}{1+a_{21}\nu_{11}(\omega)}\right)\mathrm{d}\omega. \tag{A.57}$$

So, to conclude the results for the first mode:

$$R_1^{(1)}=\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\nu_{11}(\omega)\right)\mathrm{d}\omega, \tag{A.58}$$

$$R_2^{(1)}=\min\left\{\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{a_{12}\nu_{21}(\omega)}{1+\nu_{11}(\omega)}\right)\mathrm{d}\omega,\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{\nu_{21}(\omega)}{1+a_{21}\nu_{11}(\omega)}\right)\mathrm{d}\omega\right\}. \tag{A.59}$$

For the second mode, $i=2$, we wish to calculate the mutual information terms in $\tilde{\mathcal{G}}_{B_2}$. We have:

$$R_1^{(2)}=\min\left\{\lim_{n\to\infty}\frac{I(X_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}})}{n\bar{\lambda}},\lim_{n\to\infty}\frac{I(X_1^{n\bar{\lambda}};Y_2^{n\bar{\lambda}})}{n\bar{\lambda}}\right\}, \tag{A.60}$$

$$R_2^{(2)}=\lim_{n\to\infty}\frac{I(X_2^{n\bar{\lambda}};Y_2^{n\bar{\lambda}}|X_1^{n\bar{\lambda}})}{n\bar{\lambda}}, \tag{A.61}$$

where the superscript $n\bar{\lambda}$ means that we consider the complementary block $i = n\lambda, n\lambda + 1, \ldots, n$. The calculation of these mutual information rates is exactly the same as in the first mode. One obtains:

$$R_1^{(2)} = \min\left\{\frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{\nu_{12}(\omega)}{1+a_{12}\nu_{22}(\omega)}\right)\mathrm{d}\omega, \frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\frac{a_{21}\nu_{12}(\omega)}{1+\nu_{22}(\omega)}\right)\mathrm{d}\omega\right\}, \tag{A.62}$$

$$R_2^{(2)} = \frac{1}{4\pi}\int_0^{2\pi}\log\left(1+\nu_{22}(\omega)\right)\mathrm{d}\omega. \tag{A.63}$$

Thus, by time-sharing, the achievable rate region is given by:

$$\mathcal{R} = \bigcup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{(R_1, R_2) : \begin{array}{l} R_1 \le \lambda\cdot\varphi\left[\nu_{11}\right] + \bar{\lambda}\cdot\min\left\{\varphi\left[\frac{\nu_{12}}{1+a_{12}\nu_{22}}\right], \varphi\left[\frac{a_{21}\nu_{12}}{1+\nu_{22}}\right]\right\} \\ R_2 \le \bar{\lambda}\cdot\varphi\left[\nu_{22}\right] + \lambda\cdot\min\left\{\varphi\left[\frac{a_{12}\nu_{21}}{1+\nu_{11}}\right], \varphi\left[\frac{\nu_{21}}{1+a_{21}\nu_{11}}\right]\right\} \end{array}\right\}, \tag{A.64}$$

where $\boldsymbol{\theta} \triangleq (\alpha, \beta, \lambda, \{r_{11,i}\}, \{r_{12,i}\}, \{r_{21,i}\}, \{r_{22,i}\})$.

## Appendix B
## Proof of Theorem 3

For completeness, we describe the coding scheme in more detail. During a fraction $\lambda$ of the transmission time, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by:

$$x_{1,i} = \rho_{x_1}x_{1,i-1} + \sigma_1 w_{1,i} - \kappa_1\sigma_1 w_{1,i-1},\ \ i = 1, 2, \ldots, n\lambda - 1 \tag{B.1a}$$

$$x_{2,i} = \rho_{x_2}x_{2,i-1} + \sigma_2 w_{2,i} - \kappa_2\sigma_2 w_{2,i-1},\ \ i = 1, 2, \ldots, n\lambda - 1 \tag{B.1b}$$

where $\sigma_k^2$, for $k = 1, 2$, is defined in (14), $|\rho_{x_1}|, |\rho_{x_2}|, |\kappa_1|, |\kappa_2| < 1$, $x_{1,0}$ and $x_{2,0}$ are Gaussian random variables with zero mean and variances $\alpha P_1/\lambda$ and $\beta P_2/\lambda$, $\{w_{1,i}\}$ and $\{w_{2,i}\}$ are i.i.d. Gaussian processes with zero mean and variances $\alpha P_1/\lambda$ and $\beta P_2/\lambda$, and $0 \le \alpha, \beta \le 1$. That is, the symbols of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are distributed according to a ARMA model. During the remaining fraction $\bar{\lambda} = 1 - \lambda$ of the transmission time, the components of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by:

$$x_{1,i} = \rho_{x_1}x_{1,i-1} + \sigma_1\tilde{w}_{1,i} - \kappa_1\sigma_1\tilde{w}_{1,i-1},\ \ i = n\lambda + 1, \ldots, n \tag{B.2a}$$

$$x_{2,i} = \rho_{x_2}x_{2,i-1} + \sigma_2\tilde{w}_{2,i} - \kappa_2\sigma_2\tilde{w}_{2,i-1},\ \ i = n\lambda + 1, \ldots, n, \tag{B.2b}$$

where $x_{1,n\lambda}$ and $x_{1,n\lambda}$ are Gaussian random variables with zero mean and variances $\bar{\alpha}P_1/\bar{\lambda}$ and $\bar{\beta}P_2/\bar{\lambda}$, respectively, $\tilde{w}_{1,i}$ and $\tilde{w}_{2,i}$ are i.i.d. Gaussian processes with zero mean and variances $\bar{\alpha}P_1/\bar{\lambda}$ and $\bar{\beta}P_2/\bar{\lambda}$. Note that the two input codewords satisfy the power constraints: For the first user,

$$\frac{1}{n}\mathbb{E}\left\|\boldsymbol{x}_1\right\|^2 = \lambda\frac{\alpha P_1}{\lambda} + \bar{\lambda}\frac{\bar{\alpha}P_1}{\bar{\lambda}} = P_1, \tag{B.3}$$

where we have used the fact that for an ARMA process:

$$\text{Var}\left\{x_{1,i}\right\} = \frac{\alpha P_1}{\lambda} \cdot \frac{\sigma_1^2(1+\kappa_1^2-2+\kappa_1\rho_{x_1})}{1-\rho_{x_1}^2} = \frac{\alpha P_1}{\lambda}, \tag{B.4}$$

for $i = 1, 2, \ldots, n\lambda - 1$, and similarly for $i = n\lambda + 1, \ldots, n$. A similar argument is true also for the second user.

Regarding the decoding, we consider exactly the same modes of work as in Theorem 2 (see discussion after (A.36)). Now, given the above modes, let $R_1^{(i)}$ and $R_2^{(i)}$ be the transmission rates in mode $i$. Accordingly, by a time-sharing argument, the transmission rates of the two users are

$$(R_1, R_2) = \lambda \cdot (R_1^{(1)}, R_2^{(1)}) + \bar{\lambda} \cdot (R_1^{(2)}, R_2^{(2)}). \tag{B.5}$$

Let us now calculate $R_1^{(i)}$ and $R_2^{(i)}$, for $i = 1, 2$. For $i = 1$, we wish to calculate the mutual information terms in $\tilde{\mathcal{G}}_{B_1}$, i.e.,

$$R_1^{(1)} = \lim_{n\to\infty} \frac{I(X_1^{n\lambda}; Y_1^{n\lambda}|X_2^{n\lambda})}{n\lambda} \tag{B.6}$$

$$R_2^{(1)} = \min\left\{\lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_1^{n\lambda})}{n\lambda}, \lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_2^{n\lambda})}{n\lambda}\right\} \tag{B.7}$$

We saw in (A.44), (A.46), and (A.48), that:

$$I(X_1^{n\lambda}; Y_1^{n\lambda}|X_2^{n\lambda}) = \frac{1}{2}\log\det\left[\boldsymbol{I} + \boldsymbol{R}_{x_1}^{(1)}\right], \tag{B.8}$$

$$I(X_2^{n\lambda}; Y_1^{n\lambda}) = \frac{1}{2}\log\det\left[\boldsymbol{I} + \boldsymbol{R}_{x_1}^{(1)} + a_{12}\boldsymbol{R}_{x_2}^{(1)}\right] - \frac{1}{2}\log\det\left[\boldsymbol{I} + \boldsymbol{R}_{x_1}^{(1)}\right], \tag{B.9}$$

and

$$I(X_2^{n\lambda}; Y_2^{n\lambda}) = \frac{1}{2}\log\det\left[\boldsymbol{I} + a_{21}\boldsymbol{R}_{x_1}^{(1)} + \boldsymbol{R}_{x_2}^{(1)}\right] - \frac{1}{2}\log\det\left[\boldsymbol{I} + a_{21}\boldsymbol{R}_{x_1}^{(1)}\right]. \tag{B.10}$$

Thus, we need to calculate the limits of these terms as $n \to \infty$. We use again Szegö's theorem [19-22]. Indeed, recall that $X_1^n$ and $X_2^n$ are ARMA processes. Accordingly, (B.1) can be rewritten as:

$$\boldsymbol{A}_{n\lambda} X_1^{n\lambda} = \sigma_1 \boldsymbol{B}_{n\lambda} W_1^{n\lambda} \tag{B.11}$$

where

$$\boldsymbol{A}_{n\lambda} \triangleq \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ -\rho_{x_1} & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -\rho_{x_1} & 1 \end{pmatrix}, \tag{B.12}$$

and

$$\boldsymbol{B}_{n\lambda} \triangleq \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\kappa_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\kappa_1 & 1 \end{pmatrix}, \tag{B.13}$$

which are Toeplitz matrices. Thus, we have:

$$\sigma_1^2 \boldsymbol{B}_{n\lambda} \boldsymbol{R}_{w_1} \boldsymbol{B}_{n\lambda}^T = \boldsymbol{A}_{n\lambda} \boldsymbol{R}_{x_1}^{(1)} \boldsymbol{A}_{n\lambda}^T, \tag{B.14}$$

and whence, using the fact that $\boldsymbol{R}_{w_1} = (\alpha P_1/\lambda)\boldsymbol{I}$, we obtain

$$\boldsymbol{R}_{x_1}^{(1)} = \frac{\sigma_1^2 \alpha P_1}{\lambda} \cdot \boldsymbol{A}_{n\lambda}^{-1} \boldsymbol{B}_{n\lambda} \boldsymbol{B}_{n\lambda}^T (\boldsymbol{A}_{n\lambda}^{-1})^T. \tag{B.15}$$

As before, let $\left\{\rho_{x_1,k}^{(1)}\right\}$ be the eigenvalues of $\boldsymbol{R}_{x_1}^{(1)}$. Then, using Szegö's theorem, we can show that [19, Theorems 6.1 and 6.2]:

$$\lim_{n\to\infty} \sum_{k=0}^{n\lambda-1} F(\rho_{x_1,k}^{(1)}) = \frac{1}{2\pi} \int_0^{2\pi} F(\gamma_1^{(1)}(\omega)) \mathrm{d}\omega \tag{B.16}$$

for any continuous function $F(\cdot)$, where $\gamma_1^{(1)}(\cdot)$ is defined in (16). Thus,

$$\lim_{n\to\infty} \frac{I(X_1^{n\lambda}; Y_1^{n\lambda} | X_2^{n\lambda})}{n\lambda} = \frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \gamma_1^{(1)}(\omega)\right) \mathrm{d}\omega. \tag{B.17}$$

In a similar manner,

$$\lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_1^{n\lambda})}{n\lambda} = \frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \frac{a_{12}\gamma_2^{(1)}(\omega)}{1 + \gamma_1^{(1)}(\omega)}\right) \mathrm{d}\omega, \tag{B.18}$$

and

$$\lim_{n\to\infty} \frac{I(X_2^{n\lambda}; Y_2^{n\lambda})}{n\lambda} = \frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \frac{\gamma_2^{(1)}(\omega)}{1 + a_{21}\gamma_1^{(1)}(\omega)}\right) \mathrm{d}\omega, \tag{B.19}$$

where $\gamma_2^{(1)}(\cdot)$ is defined in (16). So, to conclude the results for the first mode:

$$R_1^{(1)} = \frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \gamma_1^{(1)}(\omega)\right) \mathrm{d}\omega, \tag{B.20}$$

$$R_2^{(1)} = \min\left\{\frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \frac{a_{12}\gamma_2^{(1)}(\omega)}{1 + \gamma_1^{(1)}(\omega)}\right) \mathrm{d}\omega, \frac{1}{4\pi} \int_0^{2\pi} \log\left(1 + \frac{\gamma_2^{(1)}(\omega)}{1 + a_{21}\gamma_1^{(1)}(\omega)}\right) \mathrm{d}\omega\right\}. \tag{B.21}$$

For the second mode, $i = 2$, we wish to calculate the mutual information terms in $\tilde{\mathcal{G}}_{B_2}$. We have:

$$R_1^{(2)} = \min\left\{\lim_{n\to\infty} \frac{I(X_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}})}{n\bar{\lambda}}, \lim_{n\to\infty} \frac{I(X_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}})}{n\bar{\lambda}}\right\}, \tag{B.22}$$

$$R_2^{(2)} = \lim_{n\to\infty} \frac{I(X_2^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}}|X_1^{n\bar{\lambda}})}{n\bar{\lambda}}. \tag{B.23}$$

The calculation of these mutual information rates is exactly the same as in the first mode. One obtains:

$$R_1^{(2)} = \min\left\{\frac{1}{4\pi}\int_0^{2\pi} \log\left(1 + \frac{\gamma_1^{(2)}(\omega)}{1 + a_{12}\gamma_2^{(2)}(\omega)}\right) \mathrm{d}\omega, \frac{1}{4\pi}\int_0^{2\pi} \log\left(1 + \frac{a_{21}\gamma_1^{(2)}(\omega)}{1 + \gamma_2^{(2)}(\omega)}\right) \mathrm{d}\omega\right\}, \tag{B.24}$$

$$R_2^{(2)} = \frac{1}{4\pi}\int_0^{2\pi} \log\left(1 + \gamma_2^{(2)}(\omega)\right) \mathrm{d}\omega, \tag{B.25}$$

where $\gamma_1^{(2)}(\cdot)$ and $\gamma_2^{(2)}(\cdot)$ are defined in (16). To conclude, an achievable rate region is given by:

$$\mathcal{R}^{(1)} = \bigcup_{\alpha,\beta,\lambda,\rho_{x_1},\rho_{x_2},\kappa_1,\kappa_2} \left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq \lambda \cdot \varphi\left[\gamma_1^{(1)}\right] + \bar{\lambda} \cdot \min\left\{\varphi\left[\frac{\gamma_1^{(2)}}{1+a_{12}\gamma_2^{(2)}}\right], \varphi\left[\frac{a_{21}\gamma_1^{(2)}}{1+\gamma_2^{(2)}}\right]\right\} \\ R_2 \leq \bar{\lambda} \cdot \varphi\left[\gamma_2^{(2)}\right] + \lambda \cdot \min\left\{\varphi\left[\frac{a_{12}\gamma_2^{(1)}}{1+\gamma_1^{(1)}}\right], \varphi\left[\frac{\gamma_2^{(1)}}{1+a_{21}\gamma_1^{(1)}}\right]\right\} \end{array} \right\}. \tag{B.26}$$

## APPENDIX C
## PROOFS OF THEOREMS 4 AND 5

***Encoding:*** As before, we consider two modes of works. During the first $\lambda$ fraction of the transmission time. Here, the components of $\boldsymbol{x}_1$ are given by:

$$x_{1,i} = \rho_{x_1} x_{1,i-1} + \sqrt{1-\rho_{x_1}^2}\, w_{1,i}, \quad i = 1, 2, \ldots, n\lambda - 1 \tag{C.1}$$

where $x_{1,0}$ is a Gaussian random variable with zero mean and variance $\alpha P_1/\lambda$, $\{w_{1,i}\}$ is i.i.d. Gaussian process with zero mean and variance $\alpha P_1/\lambda$, where $0 \leq \alpha \leq 1$. We generate $2^{nR_1}$ independent codewords $\{\boldsymbol{x}_1(i)\}$ according to the above distribution. Now regarding $\boldsymbol{x}_2$, we use rate-splitting technique, as used in HK scheme, and generate $2^{nR_2'}$ and $2^{nR_2''}$ independent codewords $\{\boldsymbol{v}_2(i)\}$ and $\{\boldsymbol{u}_2(i)\}$, respectively, where $0 \leq R_2' \leq R_2$ and $R_2'' = R_2 - R_2'$, in the following way:

$$u_{2,i} = \rho_{u_2} u_{2,i-1} + \sqrt{1-\rho_{u_2}^2}\, w_{2,i}^u, \tag{C.2a}$$

$$v_{2,i} = \rho_{v_2} v_{2,i-1} + \sqrt{1-\rho_{v_2}^2}\, w_{2,i}^v, \tag{C.2b}$$

where $|\rho_{u_2}|, |\rho_{v_2}| < 1$, $u_{2,0}$ and $v_{2,0}$ are Gaussian random variables with zero mean and variances $\beta\xi_2 P_2/\lambda$ and $\beta\bar{\xi}_2 P_2/\lambda$, respectively, and $\left\{w_{2,i}^u\right\}$ and $\left\{w_{2,i}^v\right\}$ are i.i.d. Gaussian processes with zero mean and variances $\beta\xi_2 P_2/\lambda$, and $\beta\bar{\xi}_2 P_2/\lambda$, respectively, where $0 \leq \beta, \xi_2 \leq 1$. Then, the $2^{nR_2}$ codewords $\{\boldsymbol{x}_2(i,j)\}$ are given by the element-wise addition:

$$\boldsymbol{x}_2(i,j) = \boldsymbol{u}_2(i) + \boldsymbol{v}_2(j) \tag{C.3}$$

for $i = 1, 2, \ldots, 2^{nR_2''}$ and $j = 1, 2, \ldots, 2^{nR_2'}$, which denote the indexes of the codewords.

During the remaining fraction $\bar{\lambda} = 1 - \lambda$ of the transmission time, the roles of the encoders are swapped. Specifically, the components of $\boldsymbol{x}_2$ are given by:

$$x_{2,i} = \rho_{x_2} x_{2,i-1} + \sqrt{1-\rho_{x_2}^2}\tilde{w}_{2,i}, \quad i = \lambda n + 1, \ldots, n \tag{C.4}$$

where $x_{2,n\lambda}$ is a Gaussian random variable with zero mean and variance $\bar{\beta}P_2/\bar{\lambda}$, $\{\tilde{w}_{2,i}\}$ is i.i.d. Gaussian process with zero mean and variance $\bar{\beta}P_2/\bar{\lambda}$, where $0 \leq \beta \leq 1$. We generate $2^{nR_2}$ independent codewords $\{\boldsymbol{x}_2(i)\}$ according to the above distribution. Regarding $\boldsymbol{x}_1$, we use again rate-splitting technique, and generate $2^{nR_1'}$ and $2^{nR_1''}$ independent codewords $\{\boldsymbol{v}_1(i)\}$ and $\{\boldsymbol{u}_1(i)\}$, respectively, where $0 \leq R_1' \leq R_1$ and $R_1'' = R_1 - R_1'$, in the following way:

$$u_{1,i} = \rho_{u_1} u_{1,i-1} + \sqrt{1-\rho_{u_1}^2} w_{1,i}^u, \tag{C.5a}$$

$$v_{1,i} = \rho_{v_1} v_{1,i-1} + \sqrt{1-\rho_{v_1}^2} w_{1,i}^v, \tag{C.5b}$$

where $|\rho_{u_1}|, |\rho_{v_1}| < 1$, $u_{1,\lambda n}$ and $v_{1,\lambda n}$ are Gaussian random variables with zero mean and variances $\bar{\alpha}\xi_1 P_1/\bar{\lambda}$ and $\bar{\alpha}\bar{\xi}_1 P_1/\bar{\lambda}$, respectively, and $\left\{w_{1,i}^u\right\}$ and $\left\{w_{1,i}^v\right\}$ are i.i.d. Gaussian processes with zero mean and variances $\bar{\alpha}\xi_1 P_1/\bar{\lambda}$ and $\bar{\alpha}\bar{\xi}_1 P_1/\bar{\lambda}$, respectively, where $0 \leq \xi_1 \leq 1$. Then, the $2^{nR_1}$ codewords $\{\boldsymbol{x}_1(i,j)\}$ are given by the element-wise addition:

$$\boldsymbol{x}_1(i,j) = \boldsymbol{u}_1(i) + \boldsymbol{v}_1(j) \tag{C.6}$$

for $i = 1, 2, \ldots, 2^{nR_1''}$ and $j = 1, 2, \ldots, 2^{nR_1'}$. Finally, note that the two input codewords satisfy the power constraints:

$$\frac{1}{n}\mathbb{E}\left\|\boldsymbol{x}_1\right\|^2 = \lambda\frac{\alpha P_1}{\lambda} + \bar{\lambda}\left[\frac{\bar{\alpha}\xi_1 P_1}{\bar{\lambda}} + \frac{\bar{\alpha}\bar{\xi}_1 P_1}{\bar{\lambda}}\right] = P_1, \tag{C.7}$$

and similarly for the second user. As described in Subsection II-B, we consider two decoding schemes (see, discussion after (41)).

***Analysis of Decoding #1:***

Using similar methods to analyze the probability of error, as used in Appendix A, it can be readily shown that under decoding strategy #1, the following is achievable during the first mode:

$$R_1^{(1)} \leq \lim_{n\to\infty}\frac{1}{n\lambda}I(X_1^{n\lambda}; Y_1^{n\lambda}|V_2^{n\lambda}), \tag{C.8a}$$

$$R_2' \leq \lim_{n\to\infty}\frac{1}{n\lambda}I(V_2^{n\lambda}; Y_1^{n\lambda}), \tag{C.8b}$$

$$R_2' \leq \lim_{n\to\infty}\frac{1}{n\lambda}I(V_2^{n\lambda}; Y_2^{n\lambda}|U_2^{n\lambda}), \tag{C.8c}$$

$$R_2'' \leq \lim_{n\to\infty}\frac{1}{n\lambda}I(U_2^{n\lambda}; Y_2^{n\lambda}|V_2^{n\lambda}), \tag{C.8d}$$

$$R_2^{(1)} = R_2' + R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}, V_2^{n\lambda}; Y_2^{n\lambda}). \tag{C.8e}$$

Accordingly, for the second mode, the following is achievable:

$$R_2^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(X_2^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}), \tag{C.9a}$$

$$R_1' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}}), \tag{C.9b}$$

$$R_1' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | U_1^{n\bar{\lambda}}), \tag{C.9c}$$

$$R_1'' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}), \tag{C.9d}$$

$$R_1^{(2)} = R_1' + R_1'' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}}). \tag{C.9e}$$

Indeed, for the first mode, for example, we see that receiver 1 first decodes $V_2$, and thus $R_2'$ is bounded by (C.8b). Then he uses his estimate as side information to decode $X_1$, and thus $R_1$ is bounded by (C.8a). Receiver 2, on the other hand, simultaneously decodes $U_2$ and $V_2$, as in decoding for the MAC, and thus we have (C.8c)-(C.8e). Before we continue, we simplify the above regions by eliminating the virtual rates $R_1'$, $R_1''$, $R_2'$, and $R_2''$. We do that by applying Fourier-Motzkin algorithm [1, Appendix D]. For the first mode, set $R_2' = R_2^{(1)} - R_2''$, and eliminate $R_2'$ from (C.8a)-(C.8e), to obtain:

$$R_1^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(X_1^{n\lambda}; Y_1^{n\lambda} | V_2^{n\lambda}), \tag{C.10a}$$

$$R_2^{(1)} - R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(V_2^{n\lambda}; Y_1^{n\lambda}), \tag{C.10b}$$

$$R_2^{(1)} - R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(V_2^{n\lambda}; Y_2^{n\lambda} | U_2^{n\lambda}), \tag{C.10c}$$

$$R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}), \tag{C.10d}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}, V_2^{n\lambda}; Y_2^{n\lambda}). \tag{C.10e}$$

Since (C.10a) and (C.10e) are independent of $R_2''$, we ignore them. Now, collect the inequalities including $R_2''$ with positive coefficients to obtain:

$$R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}), \tag{C.11}$$

and with negative coefficients to obtain:

$$R_2^{(1)} - R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(V_2^{n\lambda}; Y_1^{n\lambda}), \tag{C.12a}$$

$$R_2^{(1)} - R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(V_2^{n\lambda}; Y_2^{n\lambda} | U_2^{n\lambda}). \tag{C.12b}$$

Next, eliminate $R_2''$ by adding (C.11) to (C.12a) and (C.12b) to obtain the inequalities not including $R_2''$:

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(V_2^{n\lambda}; Y_1^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right] \tag{C.13a}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda}\left[I(V_2^{n\lambda};Y_2^{n\lambda}|U_2^{n\lambda}) + I(U_2^{n\lambda};Y_2^{n\lambda}|V_2^{n\lambda})\right]. \tag{C.13b}$$

Thus, we obtain that for the first mode, the following is achievable:

$$R_1^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(X_1^{n\lambda};Y_1^{n\lambda}|V_2^{n\lambda}), \tag{C.14a}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda}\left[I(V_2^{n\lambda};Y_1^{n\lambda}) + I(U_2^{n\lambda};Y_2^{n\lambda}|V_2^{n\lambda})\right], \tag{C.14b}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda}\left[I(V_2^{n\lambda};Y_2^{n\lambda}|U_2^{n\lambda}) + I(U_2^{n\lambda};Y_2^{n\lambda}|V_2^{n\lambda})\right], \tag{C.14c}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}, V_2^{n\lambda};Y_2^{n\lambda}). \tag{C.14d}$$

Accordingly, for the second mode, the following is achievable:

$$R_2^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(X_2^{n\bar{\lambda}};Y_2^{n\bar{\lambda}}|V_1^{n\bar{\lambda}}), \tag{C.15a}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}}\left[I(V_1^{n\bar{\lambda}};Y_2^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|V_1^{n\bar{\lambda}})\right], \tag{C.15b}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}}\left[I(V_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|U_1^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|V_1^{n\bar{\lambda}})\right], \tag{C.15c}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}, V_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}). \tag{C.15d}$$

As before, by a time-sharing argument, the transmission rates of the two users are

$$(R_1, R_2) = \lambda \cdot (R_1^{(1)}, R_2^{(1)}) + \bar{\lambda} \cdot (R_1^{(2)}, R_2^{(2)}). \tag{C.16}$$

Finally, we evaluate each of the mutual information terms in (C.14) and (C.15). Let:

$$\gamma_{x_1}(\omega) \triangleq \frac{(1-\rho_{x_1}^2)\alpha P_1}{\lambda\,|1-\rho_{x_1}e^{j\omega}|^2} \tag{C.17a}$$

$$\gamma_{x_2}(\omega) \triangleq \frac{(1-\rho_{x_2}^2)\bar{\beta} P_2}{\bar{\lambda}\,|1-\rho_{x_2}e^{j\omega}|^2} \tag{C.17b}$$

$$\gamma_{u_1}(\omega) \triangleq \frac{(1-\rho_{u_1}^2)\xi_1\bar{\alpha} P_1}{\bar{\lambda}\,|1-\rho_{u_1}e^{j\omega}|^2} \tag{C.17c}$$

$$\gamma_{v_1}(\omega) \triangleq \frac{(1-\rho_{v_1}^2)\bar{\xi}_1\bar{\alpha} P_1}{\bar{\lambda}\,|1-\rho_{v_1}e^{j\omega}|^2} \tag{C.17d}$$

$$\gamma_{u_2}(\omega) \triangleq \frac{(1-\rho_{u_2}^2)\xi_2\beta P_2}{\lambda\,|1-\rho_{u_2}e^{j\omega}|^2} \tag{C.17e}$$

$$\gamma_{v_2}(\omega) \triangleq \frac{(1-\rho_{v_2}^2)\bar{\xi}_2\beta P_2}{\lambda\,|1-\rho_{v_2}e^{j\omega}|^2}. \tag{C.17f}$$

As before, using Szegö's theorem we obtain that:

$$\lim_{n\to\infty} \frac{1}{n\lambda} I(X_1^{n\lambda};Y_1^{n\lambda}|V_2^{n\lambda}) = \varphi\left[\frac{\gamma_{x_1}}{1+a_{12}\gamma_{u_2}}\right], \tag{C.18a}$$

$$\lim_{n\to\infty}\frac{1}{n\lambda}\left[I(V_2^{n\lambda};Y_1^{n\lambda})+I(U_2^{n\lambda};Y_2^{n\lambda}|V_2^{n\lambda})\right]=\varphi\left[\frac{a_{12}\gamma_{v_2}}{1+\gamma_{x_1}+a_{12}\gamma_{u_2}}\right]+\varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right], \tag{C.18b}$$

$$\lim_{n\to\infty}\frac{1}{n\lambda}\left[I(V_2^{n\lambda};Y_2^{n\lambda}|U_2^{n\lambda})+I(U_2^{n\lambda};Y_2^{n\lambda}|V_2^{n\lambda})\right]=\varphi\left[\frac{\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right]+\varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right], \tag{C.18c}$$

$$\lim_{n\to\infty}\frac{1}{n\lambda}I(U_2^{n\lambda},V_2^{n\lambda};Y_2^{n\lambda})=\varphi\left[\frac{\gamma_{u_2}+\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right], \tag{C.18d}$$

and

$$\lim_{n\to\infty}\frac{1}{n\bar{\lambda}}I(X_2^{n\bar{\lambda}};Y_2^{n\bar{\lambda}}|V_1^{n\bar{\lambda}})=\varphi\left[\frac{\gamma_{x_2}}{1+a_{21}\gamma_{u_1}}\right], \tag{C.19a}$$

$$\lim_{n\to\infty}\frac{1}{n\bar{\lambda}}\left[I(V_1^{n\bar{\lambda}};Y_2^{n\bar{\lambda}})+I(U_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|V_1^{n\bar{\lambda}})\right]=\varphi\left[\frac{a_{21}\gamma_{v_1}}{1+\gamma_{x_2}+a_{21}\gamma_{u_1}}\right]+\varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right], \tag{C.19b}$$

$$\lim_{n\to\infty}\frac{1}{n\bar{\lambda}}\left[I(V_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|U_1^{n\bar{\lambda}})+I(U_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}}|V_1^{n\bar{\lambda}})\right]=\varphi\left[\frac{\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right]+\varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right], \tag{C.19c}$$

$$\lim_{n\to\infty}\frac{1}{n\bar{\lambda}}I(U_1^{n\bar{\lambda}},V_1^{n\bar{\lambda}};Y_1^{n\bar{\lambda}})=\varphi\left[\frac{\gamma_{u_1}+\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right]. \tag{C.19d}$$

Thus, to summarize our results, our new achievable rate region is given by:

$$\mathcal{R}^{(2)}=\bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}}\left\{(R_1,R_2):\begin{array}{c}R_1\le R_1(\boldsymbol{\xi})\\ R_2\le R_2(\boldsymbol{\xi})\end{array}\right\} \tag{C.20}$$

where

$$\begin{aligned}R_1(\boldsymbol{\xi})\triangleq\lambda\cdot\varphi\left[\frac{\gamma_{x_1}}{1+a_{12}\gamma_{u_2}}\right]+\bar{\lambda}\min\Bigg\{&\varphi\left[\frac{\gamma_{u_1}+\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right],\varphi\left[\frac{\gamma_{v_1}}{1+a_{12}\gamma_{x_2}}\right]+\varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right],\\ &\varphi\left[\frac{a_{21}\gamma_{v_1}}{1+\gamma_{x_2}+a_{21}\gamma_{u_1}}\right]+\varphi\left[\frac{\gamma_{u_1}}{1+a_{12}\gamma_{x_2}}\right]\Bigg\},\end{aligned} \tag{C.21}$$

and

$$\begin{aligned}R_2(\boldsymbol{\xi})\triangleq\bar{\lambda}\cdot\varphi\left[\frac{\gamma_{x_2}}{1+a_{21}\gamma_{u_1}}\right]+\lambda\cdot\min\Bigg\{&\varphi\left[\frac{\gamma_{u_2}+\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right],\varphi\left[\frac{\gamma_{v_2}}{1+a_{21}\gamma_{x_1}}\right]+\varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right],\\ &\varphi\left[\frac{a_{12}\gamma_{v_2}}{1+\gamma_{x_1}+a_{12}\gamma_{u_2}}\right]+\varphi\left[\frac{\gamma_{u_2}}{1+a_{21}\gamma_{x_1}}\right]\Bigg\}.\end{aligned} \tag{C.22}$$

***Analysis of Decoding #2:***

As was done in Appendix A, we can show that under decoding strategy #2, the following is achievable for the first mode:

$$R_1^{(1)}\le\lim_{n\to\infty}\frac{1}{n\lambda}I(X_1^{n\lambda};Y_1^{n\lambda}|V_2^{n\lambda}), \tag{C.23a}$$

$$R_2'\le\lim_{n\to\infty}\frac{1}{n\lambda}I(V_2^{n\lambda};Y_1^{n\lambda}|X_1^{n\lambda}), \tag{C.23b}$$

$$R_1^{(1)}+R_2'\le\lim_{n\to\infty}\frac{1}{n\lambda}I(X_1^{n\lambda},V_2^{n\lambda};Y_1^{n\lambda}), \tag{C.23c}$$

$$R_2' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(V_2^{n\lambda}; Y_2^{n\lambda} | U_2^{n\lambda}), \tag{C.23d}$$

$$R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}), \tag{C.23e}$$

$$R_2^{(1)} = R_2' + R_2'' \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}, V_2^{n\lambda}; Y_2^{n\lambda}), \tag{C.23f}$$

while for the the second mode, we have:

$$R_2^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(X_2^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}), \tag{C.24a}$$

$$R_1' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | X_2^{n\bar{\lambda}}), \tag{C.24b}$$

$$R_2^{(2)} + R_1' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(X_2^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}}), \tag{C.24c}$$

$$R_1' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | U_1^{n\bar{\lambda}}), \tag{C.24d}$$

$$R_1'' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}), \tag{C.24e}$$

$$R_1^{(2)} = R_1' + R_1'' \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}}). \tag{C.24f}$$

As can be noticed from (C.23), in the first mode, we see that now receiver 2 simultaneously (contrary to decoding #1) decodes $V_1$ and $X_2$, and thus $R_1'$ and $R_2^{(2)}$ is bounded by (C.23a)-(C.23c). Receiver 1, again, simultaneously decodes $U_1$ and $V_1$, and thus we have (C.23d)-(C.23f). As was done for decoding #1, we can simplify the above regions, by eliminating the virtual rates $R_1'$, $R_1''$, $R_2'$, and $R_2''$, via the Fourier-Motzkin algorithm. Eventually, one obtains for the first mode that the following is achievable:

$$R_1^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(X_1^{n\lambda}; Y_1^{n\lambda} | V_2^{n\lambda}), \tag{C.25a}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} I(U_2^{n\lambda}, V_2^{n\lambda}; Y_2^{n\lambda}), \tag{C.25b}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(V_2^{n\lambda}; Y_1^{n\lambda} | X_1^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right], \tag{C.25c}$$

$$R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(V_2^{n\lambda}; Y_2^{n\lambda} | U_2^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right], \tag{C.25d}$$

$$R_1^{(1)} + R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(X_1^{n\lambda}, V_2^{n\lambda}; Y_1^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right], \tag{C.25e}$$

and for the second mode:

$$R_2^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(X_2^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}), \tag{C.26a}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} I(U_1^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}}), \tag{C.26b}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} \left[ I(V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | X_2^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}) \right], \tag{C.26c}$$

$$R_1^{(2)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} \left[ I(V_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | U_1^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}) \right], \tag{C.26d}$$

$$R_1^{(1)} + R_2^{(1)} \leq \lim_{n\to\infty} \frac{1}{n\bar{\lambda}} \left[ I(X_2^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}) \right]. \tag{C.26e}$$

Using Szegö's theorem we get that (see (C.18) and (C.19)):

$$\lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(V_2^{n\lambda}; Y_1^{n\lambda} | X_1^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right] = \varphi \left[ \frac{a_{12}\gamma_{v_2}}{1 + a_{12}\gamma_{u_2}} \right] + \varphi \left[ \frac{\gamma_{u_2}}{1 + a_{21}\gamma_{x_1}} \right], \tag{C.27a}$$

$$\lim_{n\to\infty} \frac{1}{n\lambda} \left[ I(X_1^{n\lambda}, V_2^{n\lambda}; Y_1^{n\lambda}) + I(U_2^{n\lambda}; Y_2^{n\lambda} | V_2^{n\lambda}) \right] = \varphi \left[ \frac{\gamma_{x_1} + a_{12}\gamma_{v_2}}{1 + a_{12}\gamma_{u_2}} \right] + \varphi \left[ \frac{\gamma_{u_2}}{1 + a_{21}\gamma_{x_1}} \right], \tag{C.27b}$$

and

$$\lim_{n\to\infty} \frac{1}{n\bar{\lambda}} \left[ I(V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}} | X_2^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}) \right] = \varphi \left[ \frac{a_{21}\gamma_{v_1}}{1 + a_{21}\gamma_{u_1}} \right] + \varphi \left[ \frac{\gamma_{u_1}}{1 + a_{12}\gamma_{x_2}} \right], \tag{C.28a}$$

$$\lim_{n\to\infty} \frac{1}{n\bar{\lambda}} \left[ I(X_2^{n\bar{\lambda}}, V_1^{n\bar{\lambda}}; Y_2^{n\bar{\lambda}}) + I(U_1^{n\bar{\lambda}}; Y_1^{n\bar{\lambda}} | V_1^{n\bar{\lambda}}) \right] = \varphi \left[ \frac{\gamma_{x_2} + a_{21}\gamma_{v_1}}{1 + a_{21}\gamma_{u_1}} \right] + \varphi \left[ \frac{\gamma_{u_1}}{1 + a_{12}\gamma_{x_2}} \right]. \tag{C.28b}$$

Whence, we obtained the following achievable region:

$$\mathcal{R}^{(3)} = \bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \left\{ (R_1, R_2) : \begin{array}{c} R_1 \leq \bar{R}_1(\boldsymbol{\xi}) \\ R_2 \leq \bar{R}_2(\boldsymbol{\xi}) \\ R_1 + R_2 \leq R_{\text{sum}}(\boldsymbol{\xi}) \end{array} \right\} \tag{C.29}$$

where

$$\begin{aligned} \bar{R}_1(\boldsymbol{\xi}) \triangleq \lambda \cdot \varphi \left[ \frac{\gamma_{x_1}}{1 + a_{12}\gamma_{u_2}} \right] + \bar{\lambda} \min \Bigg\{ & \varphi \left[ \frac{\gamma_{u_1} + \gamma_{v_1}}{1 + a_{12}\gamma_{x_2}} \right], \varphi \left[ \frac{\gamma_{v_1}}{1 + a_{12}\gamma_{x_2}} \right] + \varphi \left[ \frac{\gamma_{u_1}}{1 + a_{12}\gamma_{x_2}} \right], \\ & \varphi \left[ \frac{a_{21}\gamma_{v_1}}{1 + a_{21}\gamma_{u_1}} \right] + \varphi \left[ \frac{\gamma_{u_1}}{1 + a_{12}\gamma_{x_2}} \right] \Bigg\}, \end{aligned} \tag{C.30}$$

and

$$\begin{aligned} \bar{R}_2(\boldsymbol{\xi}) \triangleq \bar{\lambda} \cdot \varphi \left[ \frac{\gamma_{x_2}}{1 + a_{21}\gamma_{u_1}} \right] + \bar{\lambda} \min \Bigg\{ & \varphi \left[ \frac{\gamma_{u_2} + \gamma_{v_2}}{1 + a_{21}\gamma_{x_1}} \right], \varphi \left[ \frac{\gamma_{v_2}}{1 + a_{21}\gamma_{x_1}} \right] + \varphi \left[ \frac{\gamma_{u_2}}{1 + a_{21}\gamma_{x_1}} \right], \\ & \varphi \left[ \frac{a_{12}\gamma_{v_2}}{1 + a_{12}\gamma_{u_2}} \right] + \varphi \left[ \frac{\gamma_{u_2}}{1 + a_{21}\gamma_{x_1}} \right] \Bigg\}, \end{aligned} \tag{C.31}$$

and

$$\begin{aligned} R_{\text{sum}}(\boldsymbol{\xi}) \triangleq \lambda \cdot & \left\{ \varphi \left[ \frac{\gamma_{x_1} + a_{12}\gamma_{v_2}}{1 + a_{12}\gamma_{u_2}} \right] + \varphi \left[ \frac{\gamma_{u_2}}{1 + a_{21}\gamma_{x_1}} \right] \right\} \\ & + \bar{\lambda} \cdot \left\{ \varphi \left[ \frac{\gamma_{x_2} + a_{21}\gamma_{v_1}}{1 + a_{21}\gamma_{u_1}} \right] + \varphi \left[ \frac{\gamma_{u_1}}{1 + a_{12}\gamma_{x_2}} \right] \right\}. \end{aligned} \tag{C.32}$$

## REFERENCES


[1] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[2] X. Shang and B. Chen, *Two-user Gaussian interference channels: an information theoretic point of view*. Foundations and Trends in Communications and Information Theory, vol. 10, no. 3, pp. 247378,, December 2013.

[3] I. Sason, “On achievable rate regions for the Gaussian interference channel,” *IEEE Trans. on Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, June 2004.

[4] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. on Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[5] H. Sato, “The capacity of the Gaussian interference channel under strong interference,” *IEEE Trans. on Inf. Theory*, vol. 27, no. 6, pp. 786–788, Nov. 1981.

[6] A. B. Carleial, “A case where interference does not reduce capacity,” *IEEE Trans. on Inf. Theory*, vol. 21, no. 5, pp. 569–570, Sep. 1975.

[7] R. Bustin, H. V. Poor, and S. Shamai, “The effect of maximal rate codes on the interfering message rate,” Apr. 2015. [Online]. Available: http://arxiv.org/pdf/1404.6690v4.pdf

[8] Y. Polyanskiy and Y. Wu, “Wasserstein continuity of entropy and outer bounds for interference channels,” Apr. 2015. [Online]. Available: http://arxiv.org/abs/1504.04419

[9] M. H. M. Costa, “On the Gaussian interference channel,” *IEEE Trans. on Inf. Theory*, vol. 31, no. 5, pp. 607–615, Sep. 1985.

[10] I. Sason, “On the corner points of the capacity region of a two-user Gaussian interference channel,” *IEEE Trans. on Inf. Theory*, vol. 61, no. 7, pp. 3682–3697, July 2015.

[11] C. E. Shannon, “Two-way communication channels,” in *in Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1. Berkeley, CA: Univ. California Press,, 1961, pp. 611–644.

[12] R. Ahlswede, “The capacity region of a channel with two senders and two receivers,” *Annals of Probability*, vol. 2, no. 5, pp. 805–814, Oct. 1974.

[13] A. B. Carleial, “Interference channels,” *IEEE Trans. on Inf. Theory*, vol. IT-24, no. 7, pp. 60–70, Jan. 1978.

[14] H. Sato, “Two-user communication channels,” *IEEE Trans. on Inf. Theory*, vol. IT-23, no. 3, pp. 295–304, May 1977.

[15] H. F. Chong, M. Motani, K. G. Garg, and H. El Gamal, “On the Han-Kobayashi region for the interference channel,” *IEEE Trans. on Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, July 2008.

[16] E. C. vander Meulen, “Some reflections on the interference channel,” *Communications and Cryptography: Two Sides of One Tapestry, R. E. Blahut, D. J. Costello, U. Maurer, and T. Mittelholzer, Eds. Boston, MA: Kluwer*, pp. 409–421, 1994.

[17] C. Nair, L. Xia, and M. Yazdanpanah, “Sub-optimality of the Han-Kobayashi achievable region for interference channels,” *submitted to ISIT 2015,* Feb. 2015. [Online]. Available: http://arxiv.org/abs/1502.02589

[18] ——, “Sub-optimality of the Han-Kobayashi achievable region for interference channels.” ISIT 2015.

[19] M. R. Gray, *Toeplitz and Circulant Matrices: A review*. now, 2006.

[20] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. University of Calif. Press, Berkeley and Los Angeles, 1958.

[21] A. Widom, *Toeplitz Matrices*. in Studies in Real and Complex Analysis, edited by I.I. Hirschmann, Jr., MAA Studies in Mathematics, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[22] A. Bottcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*. SIAM, 2005.