Capacity Results for Interference Networks and Nested Cut-Set Bound

by

Reza Khosravi Farsani

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, a full characterization of the sum-rate capacity for *degraded* interference networks with any number of transmitters, any number of receivers, and any possible distribution of messages among transmitters and receivers is established. It is proved that a *successive decoding scheme* is sum-rate optimal for these networks. Moreover, it is shown that the transmission of only a certain subset of messages is sufficient to achieve the sum-rate capacity for such networks. Algorithms are presented to determine this subset of messages explicitly. The sum-rate expression for the degraded networks is then used to derive a unified outer bound on the sum-rate capacity of arbitrary (non-degraded) interference networks. Several variations of degraded networks are identified for which the derived outer bound is sum-rate optimal. Specifically, noisy interference regimes are derived for certain classes of multi-user/multi-message large interference networks. Also, network scenarios are identified where the incorporation of both successive decoding and treating interference as noise achieves their sum-rate capacity.

Next, by taking insight from the results for degraded networks, an extension to the standard cut-set bound for general communication networks is presented which is referred to as *nested cut-set bound*. This bound is derived by applying a series of cuts in a nested configuration to the network first and then bounding the information rate that flows through the cuts. The key idea for bounding step is indeed to impose a degraded arrangement among the receivers corresponding to the cuts. Therefore, the bound is in fact a generalization of the outer bound for interference networks: here cooperative relaying nodes are introduced into the problem as well but the proof style for the derivation of the outer bound remains the same. The nested cut-set bound, which uniformly holds for all general communication networks of arbitrary large sizes where any subset of nodes may cooperatively communicate to any other subset of them, is indeed tighter than the cut-set bound for networks with more than one receiver. Moreover, it includes the generalized cut-set bound for deterministic networks reported by Shomorony and Avestimehr which was originally a special case of the outer bound established for the interference networks by the author (2012).

Finally, capacity bounds for the two-user interference channel with cooperative receivers via conferencing links of finite capacities are investigated. The capacity results known for this communication scenario are limited to a very few special cases of the one-sided channel. One of the major challenges in analyzing such cooperative networks is how to establish efficient capacity outer bounds for them. In this thesis, by applying new techniques, novel capacity outer bounds are presented for the interference channels with conferencing users. Using the outer bounds, several new capacity results are proved for interesting channels with unidirectional cooperation in strong and mixed interference regimes. A fact is that the conferencing link (between receivers) may be utilized to provide one receiver with information about its corresponding signal or its non-corresponding signal (interference signal). As an interesting consequence, it is demonstrated that both strategies can be helpful to achieve the capacity of the channel. Lastly, for the case of Gaussian interference channel with conferencing receivers, it is argued that our outer bound is strictly tighter than the previous one derived by Wang and Tse.

Acknowledgements

I need to deeply appreciate my supervisor Prof. Amir K. Khandani whose great support during my graduate studies at the University of Waterloo made this research possible.

Also, I would like to appreciate Prof. Ravi R. Mazumdar and Prof. Patrick Mitran who accepted to read and review this dissertation.

To My Parents

Table of Contents

List of Figures

Chapter 1
Introduction

Fundamental capacity limits have been widely explored for various scenarios of multi-user communication networks; however, our understanding regarding the nature of information flow in large multi-message networks is still very limited [1]. A full characterization of the capacity region for even most of simple network topologies, such as the two-user classical Interference Channel (IC), is unknown.

A major category of communication networks in which each node is either a transmitter or a receiver (without any interactive/relay node) is called *single-hop networks*. These networks are also referred to as *interference networks* [1] because the essential feature which uniformly appears in all of them is the interference element. The interference networks are useful models for a broad range of practical communication scenarios, specifically for wireless networks. A detailed review of the existing literature that deals with capacity bounds for interference networks is given in [1]. Moreover, interesting similarities in the derivation of capacity bounds for the basic building blocks of the interference networks, including the Broadcast Channel (BC), the classical IC, and the Cognitive Radio Channel (CRC), are discussed in [1, 2] where some new capacity results are proved for the IC and CRC as well.

In this thesis, in Chapter 2, we study the behavior of information flow in *degraded interference networks*. Two types of degradedness are commonly recognized: the *physical* and the *stochastical* degradedness. The physically degraded interference networks are those for which the receivers can be arranged in a successive order from stronger to weaker such that the signal with respect to each receiver is statistically independent of the input signals conditioned on the signal of a stronger receiver. The stochastically degraded networks are those for which there exists an equivalent physically degraded network, where the equivalency of two networks implies that the marginal distributions of the transition probability function for both networks are identical. For interference networks, since the two equivalent networks have the same capacity region, it is not required to make a distinction between the two types of degradedness; hence, we generally refer to both of them as degraded networks.

As a main theorem of this thesis, we provide a full characterization for the sum-rate capacity of degraded interference networks with any arbitrary topologies, i.e., with any number of transmitters, any number of receivers, and any distribution of messages among transmitters and receivers. First, we establish an outer bound for the sum-rate capacity of the networks. This is indeed derived by subtle techniques for calculation of mutual information functions towards establishing a single-letter bound using the degradedness relations among the outputs. We then provide an interesting coding scheme to achieve this outer bound which yields the exact sum-rate capacity for the degraded networks. We show that for all degraded networks, a successive decoding scheme is sum-rate optimal. In this decoding strategy, each receiver decodes its messages as well as all the messages corresponding to the receivers weaker than itself in a successive order from weaker receivers to the stronger ones till its own messages. Moreover, we show that to achieve the sum-rate capacity for degraded networks, only a carefully picked subset of messages is required to be considered for the transmission scheme. To this end, we present an order to expand the messages of a given arbitrary network over certain directed graphs which we call them as *plan of messages*. The plan of messages clearly depicts both the multiple access and broadcast natures included in a given interference network. These graphical tools indeed play a central role in describing the behavior of information flow in large multi-message networks as discussed in details in [3, 4]. Using the plan of messages, we explicitly determine those messages which are required to be considered for the transmission scheme to achieve the sum-rate capacity for degraded networks. Also, we provide examples to clarify our results.

In Chapter 3, we next make use of the sum-rate capacity result for the degraded networks to establish a unified outer bound on the sum-rate capacity of general non-degraded interference networks. The idea is to enhance a non-degraded network with artificial outputs to obtain a degraded network whose capacity region includes that of the original network as a subset. Therefore, the sum-rate capacity of the artificial degraded network would be an outer bound on the sum-rate capacity of the original non-degraded network as well. By using of the derived outer bound, we obtain the sum-rate capacity for several variations of degraded networks. Specifically, we introduce new network scenarios such as *Generalized Z-Interference Networks* and *Many-to-One Interference Networks* and identify noisy interference regimes for them. In this regime, treating interference as noise is sum-rate optimal. Also, for the first time, we identify network topologies where the incorporation of both successive decoding and treating interference as noise schemes achieves their sum-rate capacity.

To shed light on the importance of our capacity results for general degraded networks, we need to emphasis that even for the very special case of multi-user interference channel, the sum-rate capacity of the *discrete* degraded channel was already unknown (though for the Gaussian case, it is derived in

[5] by a rather complicated approach). In fact, available capacity results for interference networks with more than two users, in particular the multi-user IC, are very limited [1]. The majority of research on these networks has been devoted to study the *degrees of freedom region* by interference alignment techniques [6, 7]. In general, our results provide a deep understanding regarding the nature of information flow in general single-hop communication networks, specifically, those with degraded structures. These results are also very important from the viewpoint of practical applications because they do not depend on the network topology and one can apply them to a broad range of practical communication systems. Such general results are indeed rare in network information theory [1, 8].

As a second major contribution of this thesis, by taking insight from the results for degraded networks, in Chapter 5, we present an extension to the standard cut-set bound [8] for general communication networks which we refer to as *nested cut-set bound*. This bound is derived by applying a series of cuts in a nested configuration to the network first and then bounding the information rate that flows through the cuts. The key idea for bounding step is indeed to impose a degraded arrangement among the receivers corresponding to the cuts. This idea enables us to employ the proof technique of the outer bound for degraded networks to establish a general outer bound on the capacity region of all communication networks. Therefore, the bound is in fact a generalization of our outer bound for interference networks: here cooperative relaying nodes are introduced into the problem as well but the proof style for the derivation of the outer bound remains the same. The nested cut-set bound, which uniformly holds for all general communication networks of arbitrary large sizes where any subset of nodes may cooperatively communicate to any other subset of them, is indeed tighter than the cut-set bound for networks with more than one receiver. Moreover, it includes the generalized cut-set bound for deterministic networks reported by Shomorony & Avestimehr [9] which was originally a special case of the outer bound established for the interference networks by the first author [10, 11]. We finally illustrate the efficiency of the nested cut-set bound by some examples including the well-known interference-relay channel.

Finally, in Chapter 6, we study capacity bounds for the two-user IC with cooperative receivers. The feasibility of cooperation among different users that allows them to exchange information is an important feature of many practical wireless systems. User cooperation has been shown to be a crucial way of improving performance of communication networks [12]. Specifically, it is an effective way to mitigate the interference in networks [13].

One of significant ways to set up cooperation among transmitters/receivers in a communication network is the use of conferencing links of finite capacities. In particular, modern cellular systems typically rely on some high capacity direct links between base-stations. Such configurations fall under the umbrella of channels with conferencing transmitters and/or conferencing receivers. Cooperation via conferencing links was first studied by Willems [14] for a Multiple Access Channel (MAC). Willems characterized the capacity region of the two-user MAC with conferencing transmitters. In the past decade, various communication scenarios with conferencing transmitters/receivers have been studied in network information theory [15-26]. In this thesis, we consider the two-user IC with conferencing receivers. This means, two clients, forming a two-user IC, send data to their respective base-stations, and the two base-stations are connected through links of given capacities. This scenario has been previously considered in several papers. Specifically, the capacity region of the Gaussian IC with conferencing receivers was established in [20] to within a constant gap. Other works in this regard include [21-26]. Despite considerable work on the cooperative interference channels with conferencing links, up to our knowledge, the available capacity results are limited to a very few special cases of the one-sided IC with unidirectional conferencing between receivers [25, 26]. In fact, the capacity of the two-user fully-connected IC with conferencing users was not previously known even for any of the special cases where the capacity is known for the IC without cooperation, for example, the strong interference channel [27].

Indeed, a major challenge to analyze cooperative networks in general and the IC with conferencing users in specific is how to establish efficient capacity outer bounds. In literature, generally there exist three types of outer bounds for cooperative interference networks: cut-set bounds, Sato type outer bounds [28, 29], and genie-aided outer bounds [19, 20]. The cut-set bounds and Sato type outer bounds are usually insufficient to derive capacity results or even to establish capacity to within a constant gap (for Gaussian channels). The outer bounds derived by genie-aided techniques, similar to [19, 20], are useful to establish constant-gap results for Gaussian channels; however, they are still insufficient to derive exact capacity results. In this thesis, we present a novel outer bound for the two-user IC with conferencing receivers. The derivation of our outer bound is indeed involved in subtle applications of the Csiszar-Korner identity [30] for manipulating multi-letter mutual information functions to establish consistent and well-formed single-letter constraints on the communication rates. In fact, we derive our bound by extending the constraints of the outer bound established in [2] for the IC without cooperation (which was shown to be useful to derive several capacity results) to the conferencing settings as well as presenting constraints with new structures. Using our outer bound, we prove several new capacity results for the two-user (fully-connected) IC with conferencing users for both discrete and Gaussian cases. In particular, we derive four capacity results for interesting channels with unidirectional cooperation in mixed and strong interference regimes. It is a fact that a conferencing link (between receivers) may be utilized to provide one receiver with information about its corresponding signal or its non-corresponding signal (interference). As a remarkable consequence, we demonstrate that both strategies can be helpful to achieve capacity for the IC with conferencing receivers.

Lastly, for the case of Gaussian IC, we show that the derived outer bound can be made tighter by introducing additional constraints which are derived by applying genie-aided techniques as well. As a result, we obtain a new outer bound for the Gaussian IC with conferencing receivers, which can be mathematically shown that is strictly tighter than the previous one obtained by Wang and Tse [20]. The results are illustrated by simulations.

Chapter 2
Preliminaries and Definitions

Throughout this thesis, we use the following notations: Random Variables (RVs) are denoted by upper case letters (e.g. X) and lower case letters are used to show their realization (e.g. x). The range set of X is represented by $\mathcal{X}$. The Probability Distribution Function (PDF) of X is denoted by $P_X(x)$, and the conditional PDF of X given Y is denoted by $P_{X|Y}(x|y)$; also, when representing PDFs, the arguments are sometimes omitted for brevity. The probability of the event A is expressed by $Pr(A)$. The notations $\mathbb{E}[.]$, $|.|$, $\|.\|$ stand for the expectation operator, absolute value, and cardinality, respectively. The set of nonnegative real numbers is denoted by $\mathbb{R}_+$. The notation $[1:K]$, where K is a positive integer, represents the set $\{1,\dots,K\}$. The function $\psi(x)$ is defined as: $\psi(x) \equiv \frac{1}{2}\log(1+x)$, for $x \in \mathbb{R}_+$.

Definition: *Let $A = \{A_1,\dots,A_K\}$ be an arbitrary indexed set with K elements, where $K \in \mathbb{N}$. Let Ω be an arbitrary subset of A, i.e., $\Omega \subseteq A$. The identification of the set Ω, denoted by $\underline{id}_\Omega$, is defined as follows:*

$$\underline{id}_\Omega \triangleq \{l \in [1:K] :\ A_l \in \Omega\} \tag{1}$$

2.1 General Communication Networks

A general discrete memoryless communication network with N communicating nodes denoted by $\{\mathcal{X}_1,\dots,\mathcal{X}_N,\mathcal{Y}_1,\dots,\mathcal{Y}_N,\mathbb{P}_{Y_1\dots Y_N|X_1\dots X_N}\}$ is a network which is organized by N input-output (transmitter-receiver) alphabet pairs $\{\mathcal{X}_i,\mathcal{Y}_i\}_{i=1}^n$ and a transition probability function $\mathbb{P}_{Y_1\dots Y_N|X_1\dots X_N}(y_1,\dots,y_N|x_1,\dots,x_N)$ that describes the relation between the inputs and outputs of the network. Thus, each input-output alphabet pair is corresponding to a specific communicating node as shown in Fig. 1.

Encoding and decoding scheme: For the general communication network of Fig. 1, a length-n code $\mathfrak{C}^n(R_1,\dots,R_K)$ is defined as follows: Let $\mathcal{M}_l = [1:2^{nR_l}]$ for $l = 1,\dots,K$. Consider the independent random variables $M_1,\dots,M_K$ with the range sets $\mathcal{M}_1,\dots,\mathcal{M}_K$, respectively. The messages $M_1,\dots,M_K$ are intended to be communicated over the network. Let $\mathfrak{M} \triangleq \{\mathcal{M}_1,\dots,\mathcal{M}_K\}$ and $\mathbb{M} \triangleq \{M_1,\dots,M_K\}$. Also, let $\mathcal{M}_{X_1},\dots,\mathcal{M}_{X_N}$ and $\mathcal{M}_{Y_1},\dots,\mathcal{M}_{Y_N}$ be nonempty arbitrary subsets of $\mathfrak{M}$ such that:

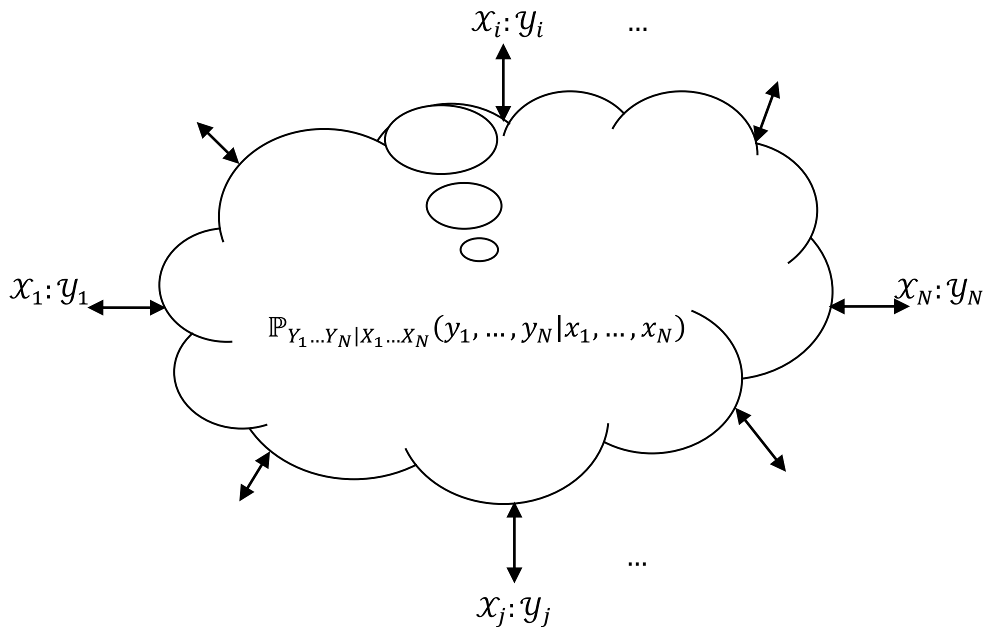

Figure 1. General Communication Netwrok.

$$\bigcup_{i \in [1:N]} \mathcal{M}_{X_i} = \bigcup_{i \in [1:N]} \mathcal{M}_{Y_i} = \mathfrak{M}, \qquad \mathcal{M}_{X_i} \cap \mathcal{M}_{Y_i} = \emptyset, \qquad i = 1, \dots, N \tag{2}$$

Define:

$$\begin{cases} \mathbb{M}_{X_i} \triangleq \left\{ M_l : \; l \in \underline{id}_{\mathcal{M}_{X_i}} \right\}, & i \in [1:N] \\ \mathbb{M}_{Y_i} \triangleq \left\{ M_l : \; l \in \underline{id}_{\mathcal{M}_{Y_i}} \right\}, & i \in [1:N] \end{cases} \tag{3}$$

The i^{th} node, $i \in [1:N]$, is designed to encode the messages $\mathbb{M}_{X_i}$ and transmit them over the network as well as to decode the messages $\mathbb{M}_{Y_i}$. To this end, the i^{th} node is associated with a set of encoding functions $\{\mathfrak{E}_{i,t}\}_{t=1}^{n}$ and a decoder function $\mathfrak{D}_i$, which operate as follows:

$$\begin{aligned} \mathfrak{E}_{i,t}: \mathcal{M}_{X_i} \times \mathcal{Y}_i^{t-1} \to \mathcal{X}_i \;\; &\Longrightarrow \;\; X_{i,t} = \mathfrak{E}_{i,t}\left(\mathbb{M}_{X_i}, Y_i^{t-1}\right) \\ \mathfrak{D}_i: \mathcal{M}_{X_i} \times Y_i^n \to \mathcal{M}_{Y_j} \;\; &\Longrightarrow \;\; \widehat{\mathbb{M}}_{Y_i} = \mathfrak{D}_i\left(\mathbb{M}_{X_i}, Y_i^n\right) \end{aligned} \tag{4}$$

The rate of the code is the K-tuple $(R_1, \dots, R_K)$. Also, the average probability of decoding, denoted by $P_e^{\mathfrak{C}^n}$, is given below:

$$P_e^{\mathfrak{C}^n} \triangleq Pr\left(\bigcup_{i\in[1:N]} \left\{\mathfrak{D}_i\left(\mathbb{M}_{X_i}, Y_i^n\right) \neq \mathbb{M}_{Y_i}\right\}\right)$$

We need to indicate that the above scenario indeed represents a general communication network in which *any subset of nodes may cooperatively send a message to any other subset of them*. It is assumed that the network is memoryless, i.e.,

$$P(x_1^n, \dots, x_N^n, y_1^n, \dots, y_N^n | m_1, \dots, m_K)$$
$$= \prod_{t=1}^{n}\prod_{i=1}^{N} P\left(x_{i,t}\middle| m_1, \dots, m_K, y_i^{t-1}\right) \mathbb{P}_{Y_1 \dots Y_N | X_1 \dots X_N}\left(y_{1,t}, \dots, y_{N,t}\middle| x_{1,t}, \dots, x_{N,t}\right) \tag{5}$$

A general Gaussian communication network with real-valued input and output signals is also formulated as follows:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} + \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix} \tag{6}$$

where the parameters $\left\{a_{ij}\right\}_{i,j=1,\dots,N}$ are (fixed) real-valued numbers, the RVs $\{X_i\}_{i=1,\dots,N}$ and $\{Y_i\}_{i=1,\dots,N}$ are input and output symbols, respectively, and the terms $\{Z_i\}_{i=1,\dots,N}$ are zero-mean unit-variance Gaussian noises. The i^{th} encoder is subject to an average power constraint as: $\mathbb{E}\left[X_i^2\right] \leq P_i$, where $P_i \in \mathbb{R}_+$, $i = 1, \dots, N$.

Definition: *A K-tuple rate $(R_1, \dots, R_K) \in \mathbb{R}_+^K$ is said to be achievable for the general communication network in Fig. 1 if for every $\epsilon > 0$ and for all sufficiently large n, there exists a length-n code $\mathfrak{C}^n(R_1, \dots, R_K)$ such that $P_e^{\mathfrak{C}^n} < \epsilon$.*

Definition: *The capacity region of the general communication network in Fig. 1, denoted by $\mathcal{C}^{GIN}$, is the closure of the set of all achievable K-tuple $(R_1, \dots, R_K)$.*

2.2 General Interference Networks

A part of the results of this thesis is specifically regarded to general single-hop communication networks in which every node is either a transmitter or a receiver and there is no relaying or active node. We generally refer to these networks as *interference networks* as the key feature of all of them is the interference element. Figure 2 depicts a general interference network composed of K_1 transmitters and K_2 receivers.

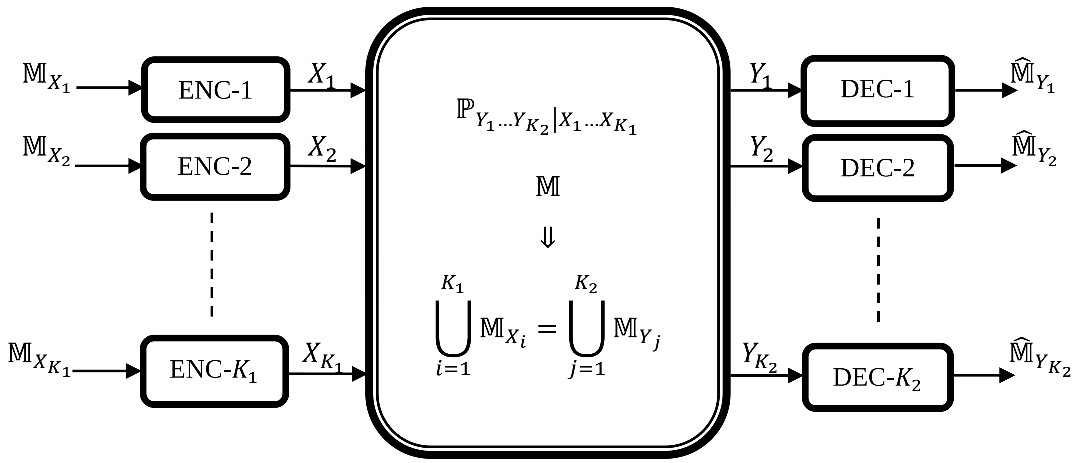

Figure 2. General Interference Netwrok.

The general interference network of Fig. 2 can be derived as a special case of the general communication network (given in Fig. 1) by setting $N = K_1 + K_2$, $Y_1 \equiv Y_2 \equiv \cdots \equiv Y_{K_1} = \emptyset$, $X_{K_1+1} \equiv X_{K_1+2} \equiv \cdots \equiv X_{K_1+K_2} \equiv \emptyset$, and then renaming the output signals $Y_{K_1+1}, Y_{K_1+2}, \dots, Y_{K_1+K_2}$ by $Y_1, Y_2, \dots, Y_{K_2}$, respectively[1]. The network transition probability function $\mathbb{P}_{Y_1 \dots Y_{K_2}|X_1 \dots X_{K_1}}(y_1, \dots, y_{K_2} | x_1, \dots, x_{K_1})$ describes the relation between the inputs and the outputs.

In particular, some of the results of the thesis are for interference networks with outputs that are statistically unrelated to some input signals. Clearly, for any interference network, *connected* and *unconnected* transmitters with respect to a given receiver are defined based on the marginal distributions of the network transition function as follows.

[1] Renaming the outputs $Y_{K_1+1}, Y_{K_1+2}, \dots, Y_{K_1+K_2}$ by $Y_1, Y_2, \dots, Y_{K_2}$, respectively, is for notation convenience only.

Definition: *Consider the general interference network in Fig. 1 with the marginal PDFs* $\mathbb{P}_{Y_1|X_1\ldots X_{K_1}},\ldots,\mathbb{P}_{Y_{K_2}|X_1\ldots X_{K_1}}$. *For the* j^{th} *receiver,* $j=1,\ldots,K_2$, *assume that* $\mathbb{P}_{Y_j|X_1\ldots X_{K_1}}$ *is as follows:*

$$\mathbb{P}_{Y_j|X_1\ldots X_{K_1}} \equiv \mathbb{P}_{Y_j|\mathbb{X}_{c\to Y_j}} \tag{7}$$

where $\mathbb{X}_{c\to Y_j}$ *is a specific subset of* $\{X_1,\ldots,X_{K_1}\}$. *In other words, the following Markov chain holds:*

$$\mathbb{X}_{c\nrightarrow Y_j} \to \mathbb{X}_{c\to Y_j} \to Y_j \tag{8}$$

where $\mathbb{X}_{c\nrightarrow Y_j} \triangleq \{X_1,\ldots,X_{K_1}\} - \mathbb{X}_{c\to Y_j}$. *In this case, the connected transmitters of the receiver* Y_j *are* $\mathbb{X}_{c\to Y_j}$ *and its unconnected transmitters are* $\mathbb{X}_{c\nrightarrow Y_j}$, $j=1,\ldots,K_2$.

2.3 Two-User Interference Channels with Conferencing Receivers

The two-user IC (as a special case of the general interference network shown in Fig. 2) is a communication scenario where two transmitters send independent messages to their corresponding users via a common media. The channel is given by the input signals $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$, the outputs $Y_1 \in \mathcal{Y}_1$ and $Y_2 \in \mathcal{Y}_2$, and the transition probability function $\mathbb{P}(y_1,y_2|x_1,x_2)$. The Gaussian channel is given in the following standard form:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + Z_1 \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + Z_2 \end{aligned} \tag{9}$$

where Z_1 and Z_2 are zero-mean unit-variance Gaussian RVs and $\mathbb{E}[X_i^2] \le P_i, i=1,2$. For the two-user Gaussian IC (9), the following notations are common in literature.

$$\mathrm{SNR}_1 := |a_{11}|^2P_1, \qquad \mathrm{SNR}_2 := |a_{22}|^2P_2, \qquad \mathrm{INR}_1 := |a_{12}|^2P_2, \qquad \mathrm{INR}_2 := |a_{21}|^2P_1 \tag{10}$$

where SNR stands for singal to noise ratio and INR stands for interference to noise ratio.

Conferencing Decoders: The two-user IC with conferencing decoders is depicted in Fig. 3. For this channel, a length-n code with L_d conferencing rounds, denoted by $\mathbb{C}^n(L_d,R_1,R_2,D_{12},D_{21})$, is described as follows.

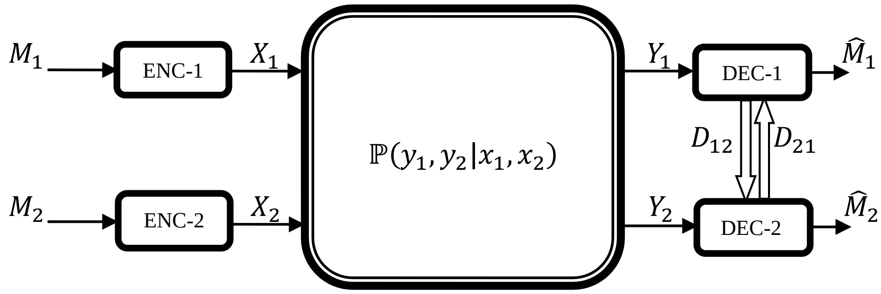

Figure 3. Two-User Interference Channel with Conferencing Decoders.

The message M_i, which is uniformly distributed over the set $\mathcal{M}_i = [1:2^{nR_i}]$, is transmitted by the i^{th} transmitter and decoded by the i^{th} receiver, $i = 1,2$. The code includes two encoder functions as:

$$\nabla_i: \mathcal{M}_i \to \mathcal{X}_i^n, \qquad X^n = \nabla_i(M_i), \qquad i \in \{1,2\}$$

Each transmitter encodes its message by the respective encoding function and sends the generated codeword over the channel. The receiver Y_i receives a sequence $Y_i^n \in \mathcal{Y}_i^n$. Before decoding process, the decoders hold a conference. The code consists of two sets of conferencing functions $\{\vartheta_{12,l}\}_{l=1}^{L_d}$ and $\{\vartheta_{21,l}\}_{l=1}^{L_d}$ with the corresponding output alphabets $\{\mathcal{V}_{12,l}\}_{l=1}^{L_d}$ and $\{\mathcal{V}_{21,l}\}_{l=1}^{L_d}$, respectively, which are described below.

$$\begin{gathered}
\vartheta_{12,l} : \mathcal{Y}_1^n \times \mathcal{V}_{21,1} \times \ldots \times \mathcal{V}_{21,l-1} \to \mathcal{V}_{12,l}, \\
V_{12,l} = \vartheta_{12,l}\left(Y_1^n, V_{21}^{l-1}\right), \\
\vartheta_{21,l} : \mathcal{Y}_2^n \times \mathcal{V}_{12,1} \times \ldots \times \mathcal{V}_{12,l-1} \to \mathcal{V}_{21,l}, \\
V_{21,l} = \vartheta_{21,l}\left(Y_2^n, V_{12}^{l-1}\right)
\end{gathered}$$

The conference is said to be (D_{12}, D_{21})-permisible if

$$\sum_{l=1}^{L_d} \log\left\|\mathcal{V}_{12,l}\right\| \le nD_{12}, \qquad \sum_{l=1}^{L_d} \log\left\|\mathcal{V}_{21,l}\right\| \le nD_{21} \tag{11}$$

The receivers exchange information by holding a (D_{12}, D_{21})-permisible conference. After the conference, the first receiver knows the sequence $V_{21}^{L_d} = \left(V_{21,1}, V_{21,2}, \ldots, V_{21,L_d}\right)$ and the second receiver knows the sequence $V_{12}^{L_d} = \left(V_{12,1}, \dots, V_{12,L_d}\right)$. The code also includes two decoder functions as follows:

$$\Delta_1\colon\ \mathcal{Y}_1^n \times \mathcal{V}_{21}^{L_d} \to \mathcal{M}_1, \qquad \widehat{M}_1 = \Delta_1\left(Y_1^n \times V_{21}^{L_d}\right)$$
$$\Delta_2\colon\ \mathcal{Y}_2^n \times \mathcal{V}_{12}^{L_d} \to \mathcal{M}_2, \qquad \widehat{M}_2 = \Delta_2\left(Y_2^n \times V_{12}^{L_d}\right)$$

Thus, each decoder decodes its message by the respective decoder function.

Lastly, the capacity region for the two-user IC with conferencing decoders is defined similar to that of general communication networks.

Chapter 3

Degraded Interference Networks: Explicit Sum-Rate Capacity[2]

In this chapter, we derive a fundamental result regarding the nature of information flow in arbitrary degraded interference networks. Clearly, we establish the sum-rate capacity for any degraded interference network with arbitrary number of transmitters, arbitrary number of receivers, and arbitrary distribution of messages among the transmitters and the receivers for both discrete and Gaussian networks. In other words, the only constraint which we impose on the underlying network is the degradedness. Let us first present a mathematical definition of degraded networks.

Definition: *Degraded Interference Networks*

The K_1-transmitter/K_2-receiver interference network shown in Fig. 2 is said to be physically degraded if there exists a permutation $\lambda(.)$ of the element of the set $[1:K_2]$ with:

$$\mathbb{P}_{Y_1\ldots Y_{K_2}|X_1\ldots X_{K_1}}\left(y_1,\ldots,y_{K_2}\middle|x_1,\ldots,x_{K_1}\right)$$
$$= \mathbb{P}_{Y_{\lambda(1)}|X_1\ldots X_{K_1}}\left(y_{\lambda(1)}\middle|x_1,\ldots,x_{K_1}\right)\mathbb{P}_{Y_{\lambda(2)}|Y_{\lambda(1)}}\left(y_{\lambda(2)}\middle|y_{\lambda(1)}\right)\ldots\mathbb{P}_{Y_{\lambda(K_2)}|Y_{\lambda(K_2-1)}}\left(y_{\lambda(K_2)}\middle|y_{\lambda(K_2-1)}\right) \tag{12}$$

Equivalently, $X_1,\ldots,X_{K_1}\to Y_{\lambda(1)}\to\cdots\to Y_{\lambda(K_2-1)}\to Y_{\lambda(K_2)}$ forms a Markov chain. In this thesis, without loss of generality, we consider the case of $\lambda(j)\equiv j$, for $j\in[1:K_2]$.

A strictly larger class of interference networks, which behave essentially similar to the physically degraded ones, is the *stochastically degraded networks* as given below.

Definition: *The K_1-transmitter/K_2-receiver interference network shown in Fig. 2 is said to be stochastically degraded if there exists a permutation $\lambda(.)$ of the element of the set $[1:K_2]$ and some transition probability functions $\widetilde{\mathbb{P}}_{Y_{\lambda(2)}|Y_{\lambda(1)}}\left(y_{\lambda(2)}\middle|y_{\lambda(1)}\right),\ldots,\widetilde{\mathbb{P}}_{Y_{\lambda(K_2)}|Y_{\lambda(K_2-1)}}\left(y_{\lambda(K_2)}\middle|y_{\lambda(K_2-1)}\right)$ such that:*

$$\mathbb{P}_{Y_j|X_1\ldots X_{K_1}}\left(y_j\middle|x_1,\ldots,x_{K_1}\right)=$$

[2] The content of this chapter was partially presented in *IEEE ISIT* 2014 [11].

$$\sum_{y_{\lambda(1)},\dots,y_{\lambda(j-1)}} \mathbb{P}_{Y_{\lambda(1)}|X_1\dots X_{K_1}}\left(y_{\lambda(1)}\middle|x_1,\dots,x_{K_1}\right)\widetilde{\mathbb{P}}_{Y_{\lambda(2)}|Y_{\lambda(1)}}\left(y_{\lambda(2)}\middle|y_{\lambda(1)}\right)\times\dots$$
$$\times\,\widetilde{\mathbb{P}}_{Y_{\lambda(j)}|Y_{\lambda(j-1)}}\left(y_{\lambda(j)}\middle|y_{\lambda(j-1)}\right),\qquad j=2,\dots,K_2 \tag{13}$$

It is known that the capacity region of all interference networks depends only on the marginal distributions of the network transition probability function. Therefore, no distinction is required to be made between stochastically and physically degradedness, and hereafter, we refer to both as the degraded interference networks.

The degraded Gaussian interference networks are also characterized in the following lemma.

Lemma 3.1: *The general Gaussian interference network is stochastically degraded provided that the network gain matrix* $\left[a_{ji}\right]_{K_2\times K_1}$ *is of rank one*.

To see a proof of the above lemma, the reader may refer to [10].

Now we are at the point to present our main theorem for degraded interference networks.

Theorem 3.1: *Consider the* K_1*-Transmitter/*K_2*-Receiver degraded interference network in (12) with the message sets* $\mathbb{M},\mathbb{M}_{X_i}, i=1,\dots,K_1$*, and* $\mathbb{M}_{Y_j}, j=1,\dots,K_2$*. The sum-rate capacity, denoted by* $\mathcal{C}_{sum}^{deg}$*, is given as follows:*

$$\mathcal{C}_{sum}^{deg}=\max_{\mathcal{P}_{sum}^{deg}}\left(I\left(\mathbb{M}_{Y_1};Y_1\middle|\mathbb{M}_{Y_2},\dots,\mathbb{M}_{Y_{K_2-1}},\mathbb{M}_{Y_{K_2}},Q\right)+\dots+I\left(\mathbb{M}_{Y_{K_2-1}};Y_{K_2-1}\middle|\mathbb{M}_{Y_{K_2}},Q\right)\right.$$
$$\left.+\,I\left(\mathbb{M}_{Y_{K_2}};Y_{K_2}\middle|Q\right)\right) \tag{14}$$

where $\mathcal{P}_{sum}^{deg}$ *denotes the set of all joint PDFs* $P_{QM_1\dots M_KX_1\dots X_{K_1}}\left(q,m_1,\dots,m_2,x_1,\dots,x_{K_1}\right)$ *satisfying:*

$$P_{QM_1\dots M_KX_1\dots X_{K_1}}=P_QP_{M_1}\dots P_{M_K}P_{X_1|\mathbb{M}_{X_1},Q}\dots P_{X_{K_1}|\mathbb{M}_{X_{K_1}},Q} \tag{15}$$

Also, the PDFs $P_{M_l}, l=1,\dots,K$*, are uniformly distributed, and* $P_{X_i|\mathbb{M}_{X_i}Q}\in\{0,1\}$ *for* $i=1,\dots,K_1$*, i.e.,* X_i *is a deterministic function of* $\left(\mathbb{M}_{X_i},Q\right)$*.*

Before presenting the proof of Theorem 3.1, first note that the sum-rate capacity expression (14) is characterized by the following parameters:

1. The RVs representing the receivers signals, i.e., $Y_j, j = 1, \dots, K_2$.
2. The auxiliary RVs representing the messages $\mathbb{M} \triangleq \{M_1, \dots, M_K\}$.
3. The parameter Q which is the time-sharing RV.

In general, the auxiliaries $M_1, \dots, M_K$ are essential for representing the sum-rate capacity of degraded networks, however, for many networks, the sum-rate capacity expression (14) can be considerably simplified. We will later discuss this issue in details.

Proof of Theorem 3.1) The achievability of (14) in fact can be derived based on a simple coding scheme. The essence of this coding scheme, which is a successive decoding algorithm, indeed can be perceived by the structure of the expression (14). Clearly, consider the following random coding of length n.

For encoding, all of the messages are encoded separately using independent codewords. Typically, any message $M_k \in \mathbb{M}$ is encoded using a codeword M_k^n generated i.i.d. based on $P_{M_k}(m_k)$. Also, a time-sharing codeword Q^n is generated independently based on $P_Q(q)$ and is revealed to all parties. Then, each transmitter $X_i \in \left\{X_1, \dots, X_{K_1}\right\}$ generates its codeword X_i^n as $X_i^n = f_i\left(\left\{(M_k)^n : M_k \in \mathbb{M}_{X_i}\right\}, Q^n\right)$, where $f_i(.)$ is an arbitrary deterministic function, and transmits it. At the receivers, the respective codewords are decoded successively. At the weakest receiver Y_{K_2}, the codewords $\left\{(M_k)^n : M_k \in \mathbb{M}_{Y_{K_2}}\right\}$ are successively decoded (in any arbitrary order). The partial sum-rate due to this step is:

$$I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2} \middle| Q\right)$$

At the receiver Y_{K_2-1}, the codewords $\left\{(M_k)^n : M_k \in \mathbb{M}_{Y_{K_2}}\right\}$ are successively decoded first similar to the receiver Y_{K_2}; this step does not introduce any new rate cost because Y_{K_2} is a degraded version of Y_{K_2-1}. Then, the codewords $\left\{(M_k)^n : M_k \in \mathbb{M}_{Y_{K_2-1}} - \mathbb{M}_{Y_{K_2}}\right\}$ are decoded again successively. The partial sum-rate due to this step would be the following:

$$I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1} \middle| \mathbb{M}_{Y_{K_2}}, Q\right)$$

This process is followed at the other receivers step by step from the weaker receivers towards the stronger ones. One can easily check that by this scheme the sum-rate (14) is achieved.

Next we present the converse proof. The derivation of the bound (14) indeed requires subtle calculations as given in the following. Also, we need to emphasis that, in Chapter 4, we make use of a similar proof style to derive the nested cut-set bound for general communication networks. Consider a code of length n with the communication rates $R_1, \dots, R_K$ for the messages $M_1, \dots, M_K$, respectively. Assume that the code has vanishing average error probability. Define the sets $\mathcal{L}_j, j = 1, \dots, K_2$, as follows:

$$\mathcal{L}_j \triangleq \mathbb{M}_{Y_j} - \left(\mathbb{M}_{Y_{j+1}} \cup \dots \cup \mathbb{M}_{Y_{K_2}}\right), \qquad j = 1, \dots, K_2 \tag{16}$$

where $\mathbb{M}_{Y_{K_2+1}} \triangleq \emptyset$. Let recall that in the following analysis, for a given subset of messages Ω, the notation $\underline{id}_{\Omega}$ denotes the identification of the set Ω, as defined in (1). Now using Fano's inequality, we can write:

$$\begin{aligned}
\sum_{l \in \underline{id}_{\mathcal{L}_{K_2}}} R_l &\leq \frac{1}{n} I\left(\mathcal{L}_{K_2}; Y_{K_2}^n\right) + \epsilon_{K_2,n} \\
&= \frac{1}{n} \sum_{t=1}^{n} I\left(\mathcal{L}_{K_2}; Y_{K_2,t} \middle| Y_{K_2}^{t-1}\right) + \epsilon_{K_2,n} \\
&\leq \frac{1}{n} \sum_{t=1}^{n} I\left(\mathcal{L}_{K_2}, Y_{K_2}^{t-1}; Y_{K_2,t}\right) + \epsilon_{K_2,n} \\
&\overset{(a)}{=} \frac{1}{n} \sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}; Y_{K_2,t}\right) + \epsilon_{K_2,n}
\end{aligned} \tag{17}$$

where $\epsilon_{K_2,n} \to 0$ as $n \to \infty$, and equality (a) holds because $\mathcal{L}_{K_2} = \mathbb{M}_{Y_{K_2}}$. Also, we have:

$$\begin{aligned}
\sum_{l \in \underline{id}_{\mathcal{L}_{K_2-1}}} R_l &\leq \frac{1}{n} I\left(\mathcal{L}_{K_2-1}; Y_{K_2-1}^n\right) + \epsilon_{K_2-1,n} \\
&\leq \frac{1}{n} I\left(\mathcal{L}_{K_2-1}; Y_{K_2-1}^n, Y_{K_2}^n, \mathbb{M}_{Y_{K_2}}\right) + \epsilon_{K_2-1,n}
\end{aligned}$$

$$\overset{(a)}{=} \frac{1}{n} I\left(\mathcal{L}_{K_2-1}; Y_{K_2-1}^n, Y_{K_2}^n \middle| \mathbb{M}_{Y_{K_2}}\right) + \epsilon_{K_2-1,n}$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathcal{L}_{K_2-1}; Y_{K_2-1,t}, Y_{K_2,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}\right) + \epsilon_{K_2-1,n}$$

$$\overset{(b)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathcal{L}_{K_2-1}; Y_{K_2-1,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}\right) + \epsilon_{K_2-1,n}$$

$$\leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathcal{L}_{K_2-1}, Y_{K_2-1}^{t-1}; Y_{K_2-1,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}\right) + \epsilon_{K_2-1,n}$$

$$\overset{(c)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}; Y_{K_2-1,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}\right) + \epsilon_{K_2-1,n} \tag{18}$$

where $\epsilon_{K_2-1,n} \to 0$ as $n \to \infty$; equality (a) holds because the elements in the set $\mathcal{L}_{K_2-1}$ are independent of those in $\mathbb{M}_{Y_{K_2}}$, equality (b) is due to degradedness property of the channel and equality (c) holds because $\left(\mathbb{M}_{Y_{K_2-1}} - \mathcal{L}_{K_2-1}\right) \subseteq \mathbb{M}_{Y_{K_2}}$. Similarly, one can derive:

$$\sum_{l \in \underline{id}_{\mathcal{L}_j}} R_l \leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_j}, Y_j^{t-1}; Y_{j,t} \middle| \mathbb{M}_{Y_{j+1}}, \dots, \mathbb{M}_{Y_{K_2}}, Y_{j+1}^{t-1}, \dots, Y_{K_2}^{t-1}\right) + \epsilon_{j,n}, \qquad j = 1, \dots, K_2 \tag{19}$$

where $\epsilon_{j,n} \to 0$ as $n \to \infty$. Define:

$$\Sigma_j \triangleq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_j}, Y_j^{t-1}; Y_{j,t} \middle| \mathbb{M}_{Y_{j+1}}, \dots, \mathbb{M}_{Y_{K_2}}, Y_{j+1}^{t-1}, \dots, Y_{K_2}^{t-1}\right) \tag{20}$$

Now we have:

$$\Sigma_{K_2} + \Sigma_{K_2-1} = \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}; Y_{K_2-1,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(Y_{K_2}^{t-1}; Y_{K_2,t} \middle| \mathbb{M}_{Y_{K_2}}\right)$$

$$+ \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}; Y_{K_2-1,t} \middle| \mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}\right)$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(Y_{K_2}^{t-1}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}, Y_{K_2}^{t-1}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right)
\end{aligned} \tag{21}
$$

where inequality (a) is due to the degradedness property of the channel. Moreover,

$$
\begin{aligned}
&\Sigma_{K_2} + \Sigma_{K_2-1} + \Sigma_{K_2-2} \\
&\overset{(a)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}, Y_{K_2-2}^{t-1}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}, Y_{K_2-2}^{t-1}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}\right) \\
&\overset{(b)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}, Y_{K_2-2}^{t-1}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}\right)
\end{aligned}
$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}, Y_{K_2-2}^{t-1}, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) \tag{22}$$

where inequality (a) is obtained from (20) and (21), and (b) is due to the degradedness property of the channel. By continuing the steps (21) and (22), one can derive:

$$\Sigma_{K_2} + \Sigma_{K_2-1} + \Sigma_{K_2-2} + \cdots + \Sigma_1$$
$$\leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) + \cdots +$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_2}; Y_{2,t}\middle|\mathbb{M}_{Y_3}, \dots, \mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_1}, Y_1^{t-1}, \dots, Y_{K_2-1}^{t-1}, Y_{K_2}^{t-1}; Y_{1,t}\middle|\mathbb{M}_{Y_2}, \dots, \mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right)$$
$$\overset{(a)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}}; Y_{K_2,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}}; Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-2}}; Y_{K_2-2,t}\middle|\mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) + \cdots +$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_2}; Y_{2,t}\middle|\mathbb{M}_{Y_3}, \dots, \mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_1}; Y_{1,t}\middle|\mathbb{M}_{Y_2}, \dots, \mathbb{M}_{Y_{K_2-1}}, \mathbb{M}_{Y_{K_2}}\right) \tag{23}$$

where equality (a) holds because:

$$\frac{1}{n}\sum_{t=1}^{n} I\left(Y_1^{t-1},\dots,Y_{K_2-1}^{t-1},Y_{K_2}^{t-1};Y_{1,t}\middle|\mathbb{M}_{Y_1},\mathbb{M}_{Y_2},\dots,\mathbb{M}_{Y_{K_2-1}},\mathbb{M}_{Y_{K_2}}\right)$$

$$=\frac{1}{n}\sum_{t=1}^{n} I\left(Y_1^{t-1},\dots,Y_{K_2-1}^{t-1},Y_{K_2}^{t-1};Y_{1,t}\middle|\mathbb{M}\right)=0 \tag{24}$$

Thus, using (23), we can obtain:

$$\begin{aligned}
\mathcal{C}_{sum}^{deg} &= \sum_{j=1}^{K_2}\left(\sum_{l\in \underline{id}_{\mathcal{L}_j}} R_l\right) \leq \sum_{j=1}^{K_2} \Sigma_j + \sum_{j=1}^{K_2}\epsilon_{j,n} \\
&\leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_1};Y_{1,t}\middle|\mathbb{M}_{Y_2},\dots,\mathbb{M}_{Y_{K_2-1}},\mathbb{M}_{Y_{K_2}}\right)+\dots+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2-1}};Y_{K_2-1,t}\middle|\mathbb{M}_{Y_{K_2}}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{Y_{K_2}};Y_{K_2,t}\right)+\sum_{j=1}^{K_2}\epsilon_{j,n}
\end{aligned} \tag{25}$$

Finally, by applying a standard time-sharing argument and also by letting n tends to infinity, we derive the outer bound (14). The proof is thus complete. ■

Theorem 3.1 establishes the sum-rate capacity for all degraded interference networks with arbitrary number of transmitters, arbitrary number of receivers, and arbitrary distribution of messages among transmitters and receivers. However, the characterization of the sum-rate capacity for degraded interference networks as given in Theorem 3.1 requires representing all messages by auxiliary random variables. In fact, many of these auxiliaries may be redundant. Clearly, to achieve the sum-rate capacity for degraded interference networks, it is always sufficient to consider only a certain subset of messages for the transmission scheme. In what follows, we present an algorithm to determine this subset of messages exactly. This algorithm indeed yields a considerably simpler characterization of the sum-rate capacity as well. To present the algorithm, we need to discuss a plan of messages as given in the following.

Plan of Messages: Consider an arbitrary interference network with the message sets $\mathbb{M},\mathbb{M}_{X_i}, i=1,\dots,K_1$, and $\mathbb{M}_{Y_j}, j=1,\dots,K_2$ as shown in Fig. 2. Each subset of transmitters sends at most one message to each subset of receivers. There exist K_1 transmitters and K_2 receivers. Therefore, we can label each message by a nonempty subset of $\{1,\dots,K_1\}$ to determine which transmitters transmit the message and also a nonempty subset of $\{1,\dots,K_2\}$ to determine to which subset of receivers the message is sent. We represent each message of $\mathbb{M}$ as M_Δ^∇, where $\Delta \subseteq \{1,\dots,K_1\}$ and $\nabla \subseteq \{1,\dots,K_2\}$. For example, $M_{\{1,2,3\}}^{\{2,4\}}$ indicates a message which is sent by Transmitters 1, 2 and 3 to Receivers 2 and 4. Now, for each $\Delta \subseteq \{1,\dots,K_1\}$ define:

$$\mathbb{M}_\Delta \triangleq \left\{M_\Delta^L \in \mathbb{M} : L \subseteq \{1,\dots,K_2\}\right\} \tag{26}$$

Using this representation, we can arrange the messages into a graph-like illustration as shown in Fig. 4. The sets $\mathbb{M}_\Delta$ are positioned on K_1 columns as follows: $\mathbb{M}_\Delta$ is positioned in the column i if and only if $\|\Delta\| = i$, $i = 1,\dots,K_1$. Then, considering column K_1 towards column 1, a set $\mathbb{M}_{\Delta_1}$ is connected to $\mathbb{M}_{\Delta_2}$ by a directed edge if and only if $\Delta_2 \subseteq \Delta_1$ and there is no nonempty $\mathbb{M}_{\Delta_3}$ so that $\Delta_2 \subseteq \Delta_3 \subseteq \Delta_1$, i.e., there is no directed path connecting $\mathbb{M}_{\Delta_1}$ to $\mathbb{M}_{\Delta_2}$. Figure 4 shows the plan of messages for a general interference network.

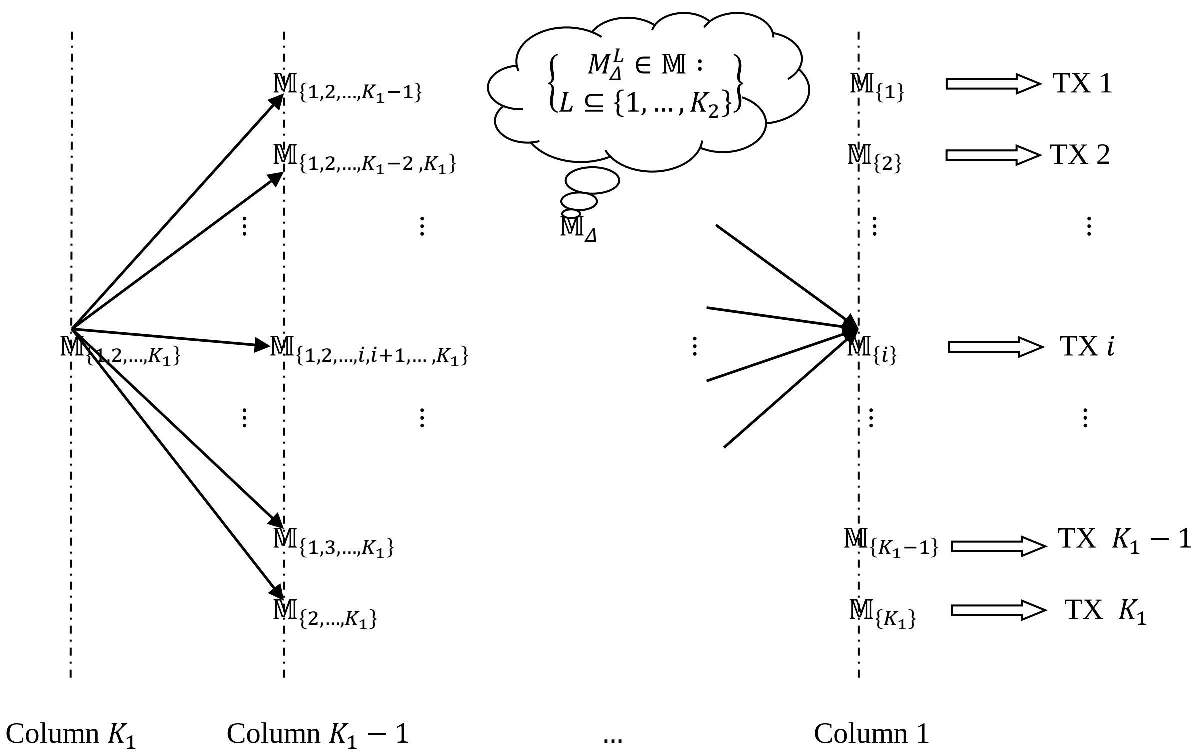

Figure 4. Plan of Messages for an arbitrary interference network.

The plan of messages indeed represents both the broadcast and multiple access features of interference networks simultaneously. Let Δ be an arbitrary nonempty subset of $\{1, \dots, K_1\}$. According to our representation, the messages $\mathbb{M}_{\Delta}$ given in (26) are broadcasted by the transmitters $X_i, i \in \Delta$ (meanwhile, no transmitter other than those in $\{X_i, i \in \Delta\}$ has access to these messages). By another view, if the network has just one receiver, it reduces to a multiple access channel and therefore each set $\mathbb{M}_{\Delta}, \Delta \subseteq \{1, \dots, K_1\}$, includes at most one message. In this case, the plan of messages indeed represents the superposition coding scheme which achieves the capacity region of the channel: any message at the beginning of an edge would be a cloud center and the message at the end of that edge would be its satellite. A more detailed discussion in this regard can be found in [10]. The plan of messages is indeed a very useful tool to describe and also to design achievability schemes for the interference networks as well (see [3, 5]).

Now using the plan of messages, we explicitly determine which messages are sufficient to be considered for the transmission scheme to achieve the sum-rate capacity of degraded networks. In what follows, we illustrate the procedure for this goal and describe the philosophy behind its steps, however, we do not provide a mathematical proof of it for brevity. A complete mathematical proof can be found in [10].

First let consider a multi-receiver broadcast channel where a transmitter sends some messages to some receivers. If the channel is degraded, it is not difficult to show that to achieve the sum-rate capacity it is sufficient to transmit only a carefully picked message and ignore the others. Inspired by this fact, one can prove that to achieve the maximum sum-rate in any degraded interference network, for every Δ, it is optimal to transmit only one of the messages in $\mathbb{M}_{\Delta}$ and ignore the others. This message is chosen as follows. Let $\mathbb{M}_{\Delta} = \left\{M_{\Delta}^{\nabla_1}, \dots, M_{\Delta}^{\nabla_{\|\mathbb{M}_{\Delta}\|}}\right\}$, where $\nabla_l \subseteq \{1, \dots, K_2\}$, $l = 1, \dots, K_2$. For any $\mathbb{M}_{\Delta}$ with $\|\mathbb{M}_{\Delta}\| \geq 1$, define:

$$\Theta_{\Delta} \triangleq \min\left\{\max \nabla : \nabla \subseteq \{1, \dots, K_2\}, M_{\Delta}^{\nabla} \in \mathbb{M}_{\Delta}\right\} = \min\left\{\max \nabla_1, \dots, \max \nabla_{\|\mathbb{M}_{\Delta}\|}\right\} \tag{27}$$

Let also $\vartheta_{\Delta} \in \{1, \dots, \|\mathbb{M}_{\Delta}\|\}$ be such that $\max \nabla_{\vartheta_{\Delta}} = \Theta_{\Delta}$. If there exist multiple choices for ϑ_{Δ}, one is selected arbitrarily. Now, to achieve the sum-rate capacity of degraded interference networks, it is sufficient to transmit only the message $M_{\Delta}^{\nabla_{\vartheta_{\Delta}}}$ out of the set $\mathbb{M}_{\Delta}$ and withdraw the others from transmitting. By another point of view, one can show that a solution to the maximization (14) is to nullify all the random variables $\mathbb{M}_\Delta - \left\{M_\Delta^{\nabla_{\vartheta_\Delta}}\right\}$ for each $\Delta \subseteq \{1, \dots, K_1\}$. Therefore, let define the sets $\widetilde{\mathbb{M}}, \widetilde{\mathbb{M}}_{X_i}, i = 1, \dots, K_1$, and $\widetilde{\mathbb{M}}_{Y_j}, j = 1, \dots, K_2$, as follows:

$$\begin{cases} \widetilde{\mathbb{M}} \triangleq \mathbb{M} - \bigcup_{\Delta \subseteq \{1,\dots,K_1\}} \left(\mathbb{M}_\Delta - \left\{M_\Delta^{\nabla_{\vartheta_\Delta}}\right\}\right) \\ \widetilde{\mathbb{M}}_{X_i} \triangleq \mathbb{M}_{X_i} - \bigcup_{\Delta \subseteq \{1,\dots,K_1\}} \left(\mathbb{M}_\Delta - \left\{M_\Delta^{\nabla_{\vartheta_\Delta}}\right\}\right) \\ \widetilde{\mathbb{M}}_{Y_j} \triangleq \mathbb{M}_{Y_j} - \bigcup_{\Delta \subseteq \{1,\dots,K_1\}} \left(\mathbb{M}_\Delta - \left\{M_\Delta^{\nabla_{\vartheta_\Delta}}\right\}\right) \end{cases} \tag{28}$$

The sum-rate capacity of a degraded interference network with the message sets $\mathbb{M}, \mathbb{M}_{X_i}, i = 1, \dots K_1$, $\mathbb{M}_{Y_j}, j = 1, \dots, K_2$, is the same as that of the network with the message sets $\widetilde{\mathbb{M}}, \widetilde{\mathbb{M}}_{X_i}, i = 1, \dots K_1$, $\widetilde{\mathbb{M}}_{Y_j}, j = 1, \dots, K_2$. Therefore, from now on, we consider the network with the new message sets given by (28). To achieve the sum-rate capacity of the degraded network, still some of the messages $\widetilde{\mathbb{M}}$ are redundant and can be canceled. First, note that if we consider the plan of messages for $\widetilde{\mathbb{M}}$, each vertices of the corresponding graph represents at most one message (by the definition (28)). For such a plan of messages, we refer to the message at the beginning of any edge as a cloud center and the message at the end of it as the satellite. For a while, let us consider a multiple access channel. As mentioned before, for this channel the plan of messages represents the superposition coding which achieves the capacity region. A fact is that to achieve the sum-rate capacity of a multiple access channel with any arbitrary distribution of messages among transmitters, it is required to transmit only a certain subset of messages. These messages are indeed those which are not satellite for any other message. A proof of this result can be found in [10]. For example, for a two-user multiple access channel with a common message between transmitters, the sum-rate capacity is achieved by transmitting only the common message at its maximum rate and ignoring the private messages. Let now return to the interference network with the message set $\widetilde{\mathbb{M}}$ as given in (28). Inspired by what stated for the multiple access channel, one can prove that to achieve the sum-rate capacity of the degraded interference network, the optimal strategy is as follows: among the messages corresponding to each receiver, those which are a satellite for the messages that should be decoded at either that receiver or some stronger receivers are canceled and only the remaining messages are considered for transmitting. Clearly, define:

$$\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}} \triangleq \widetilde{\mathbb{M}}_{Y_j} - \left(\widetilde{\mathbb{M}}_{Y_{j+1}} \cup \dots \cup \widetilde{\mathbb{M}}_{Y_{K_2}}\right), \qquad j = 1, \dots, K_2 \tag{29}$$

Note that $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}, j = 1, \dots, K_2$, constitute a partition for the set $\widetilde{\mathbb{M}}$, i.e.,

$$\widetilde{\mathbb{M}} = \bigcup_{j\in[1:K_2]} \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}, \qquad \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j_1}}} \bigcap \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j_2}}} = \emptyset, \qquad j_1, j_2 \in [1:K_2], \qquad j_1 \neq j_2 \tag{30}$$

Also, for $j = 1, \dots, K_2$, define:

$$\begin{aligned} &\mathbb{M}^{*}_{Y_j} \triangleq \left\{ M_{\Delta}^{\nabla} \in \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}} : \text{There is no } M_{\Gamma}^{L} \in \widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j+1}}} \cup \dots \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{K_2}}}\right) \text{ with } \Delta \subsetneq \Gamma \right\} \\ &\mathbb{M}^{\times}_{Y_j} \triangleq \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}} - \mathbb{M}^{*}_{Y_j} \end{aligned} \tag{31}$$

According to the definition (31), the messages $\mathbb{M}^{*}_{Y_j}$ are those of $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}$ which are not satellite for any message in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j+1}}} \cup \dots \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{K_2}}}\right)$, while $\mathbb{M}^{\times}_{Y_j}$ are those of $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}$ which are a satellite for at least one message in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j+1}}} \cup \dots \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{K_2}}}\right)$. In fact, given a certain message of $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}$ to determine whether it belongs to $\mathbb{M}^{*}_{Y_j}$ or to $\mathbb{M}^{\times}_{Y_j}$, it is required to explore among all the messages in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j+1}}} \cup \dots \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{K_2}}}\right)$. If there is no cloud center for the message (equivalently, the message is not a satellite for any other message in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{j+1}}} \cup \dots \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_{K_2}}}\right)$), it belongs to $\mathbb{M}^{*}_{Y_j}$; otherwise, it belongs to $\mathbb{M}^{\times}_{Y_j}$. By using the plan of messages, the sets $\mathbb{M}^{*}_{Y_j}$ and $\mathbb{M}^{\times}_{Y_j}$ can be readily determined by inspection as will be illustrated by examples later. To achieve the maximum sum-rate in the degraded interference networks, it is optimal to transmit only the messages $\mathbb{M}^{*}_{Y_j}$ out of the set $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}$ and ignore the others, i.e., $\mathbb{M}^{\times}_{Y_j}$, for all $j \in [1:K_2]$. In other words, if a message of $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_j}}$ has a cloud center which should be decoded at the receiver Y_j or a stronger receiver, the auxiliary random variable corresponding to that message can be nullified in the sum-capacity expression (14). Therefore, let define:

$$\begin{cases} \mathbb{M}^* \triangleq \bigcup_{j\in[1:K_2]} \mathbb{M}_{Y_j}^* \\ \mathbb{M}_{X_i}^* \triangleq \widetilde{\mathbb{M}}_{X_i} - \left(\mathbb{M}_{Y_1}^{\times} \cup \dots \cup \mathbb{M}_{Y_{K_2}}^{\times}\right), \qquad i = 1, \dots, K_1 \end{cases}$$

(32)

where $\mathbb{M}_{Y_j}^*, \mathbb{M}_{Y_j}^{\times}, j = 1, \dots, K_2$, are given by (31), and $\widetilde{\mathbb{M}}_{X_i}, i = 1, \dots, K_1$ is given by (28). The sum-rate capacity of a degraded interference network with the message sets $\mathbb{M}, \mathbb{M}_{X_i}, i = 1, \dots K_1, \mathbb{M}_{Y_j}, j = 1, \dots, K_2$, is the same as that of the network with the message sets $\mathbb{M}^*, \mathbb{M}_{X_i}^*, i = 1, \dots K_1, \mathbb{M}_{Y_j}^*, j = 1, \dots, K_2$.

We summarize the result of the above algorithm in the following theorem.

Theorem 3.2: *Consider the K_1-Transmitter/K_2-Receiver degraded interference network given in (12) with the message sets $\mathbb{M}, \mathbb{M}_{X_i}, i = 1, \dots, K_1$, and $\mathbb{M}_{Y_j}, j = 1, \dots, K_2$. The sum-rate capacity is given by:*

$$\mathcal{C}_{sum}^{deg} = \max_{\overset{*}{\mathcal{P}}_{sum}^{deg}} \Big(I\left(\mathbb{M}_{Y_1}^*; Y_1 \middle| \mathbb{M}_{Y_2}^*, \dots, \mathbb{M}_{Y_{K_2-1}}^*, \mathbb{M}_{Y_{K_2}}^*, Q\right) + \dots + I\left(\mathbb{M}_{Y_{K_2-1}}^*; Y_{K_2-1} \middle| \mathbb{M}_{Y_{K_2}}^*, Q\right) + I\left(\mathbb{M}_{Y_{K_2}}^*; Y_{K_2} \middle| Q\right)\Big)$$

(33)

where $\overset{*}{\mathcal{P}}_{sum}^{deg}$ denotes the set of all joint PDFs as follows:

$$P_Q \times \prod_{M_\Delta^\nabla \in \mathbb{M}^*} P_{M_\Delta^\nabla} \times \prod_{i\in[1:K_1]} P_{X_i|\mathbb{M}_{X_i}^* Q}$$

(34)

also, the PDFs $P_{M_\Delta^\nabla}$, $M_\Delta^\nabla \in \mathbb{M}^*$ are uniformly distributed, and $P_{X_i|\mathbb{M}_{X_i}^* Q} \in \{0,1\}$ for $i = 1, \dots, K_1$, i.e., X_i is a deterministic function of $\left(\mathbb{M}_{X_i}^*, Q\right)$. The messages $\mathbb{M}^*$ are given in (32).

Remarks:

1. The sum-rate capacity result (33) holds also for the Gaussian degraded interference networks. For these networks, one can easily solve the corresponding optimization. In fact, one can prove that the Gaussian input distributions are always optimal as will be illustrated by some examples

later. Nevertheless, providing a closed-form expression for the general case is not desirable due to the high formulation and computational complexities.

2. Recall the coding scheme given in Theorem 3.1 to achieve the sum-rate capacity of degraded networks. In that scheme, all messages are encoded separately using independent codewords. A fact is that based on the algorithm described above for determining the messages which are effective in computation of the sum-rate capacity for degraded networks, one can derive an alternative achievability scheme using the superposition coding. Briefly, considering the plan of messages for $\mathbb{M}^*$ as given in Theorem 3.2, the messages are encoded at the transmitters by superposition coding so that the message at the beginning of any edge is encoded by a cloud center codeword and the message at the end of it by a satellite codeword that is superimposed on the cloud center. For decoding, we can still use the successive decoding scheme. One can easily show that such a coding scheme does successfully work and yields the sum-rate (33).

A significant point regarding the algorithm formulated in Theorem 3.2 is that it provides us a deep insight regarding the nature of information flow in degraded networks because it explicitly determines which messages are really required to be transmitted to achieve the sum-rate capacity for these networks. In general, by using this algorithm, many of the messages (in fact auxiliary random variables) are removed from the initial sum capacity expression given in Theorem 3.1. Therefore, the maximization problem which is contained in the characterization of the sum-rate capacity is considerably resolved. Moreover, the simplicity of our characterization in Theorem 3.2 is crucial when we consider the Gaussian networks because the complexity in the explicit characterization of the sum-rate capacity for these networks grows very rapidly with the number of messages. This simplicity is also very helpful when considering the outer bound which will be derived later in Chapter 4 for the sum-rate capacity of general interference networks.

In the rest of this chapter, we provide several examples to illustrate the results of Theorem 3.1 and 3.2. Note that for all of the following examples, we assume that the network is degraded in the sense of (12).

Example 3.1: Consider a 4-transmitter/3-receiver interference network with the following message set:

$$\mathbb{M} = \left\{ \begin{array}{c} M_{\{1,2,4\}}^{\{3\}}, M_{\{1,2,4\}}^{\{1,3\}}, M_{\{1,2,4\}}^{\{2,3\}}, M_{\{1,2\}}^{\{2,3\}}, M_{\{1,2\}}^{\{1,3\}}, M_{\{3,4\}}^{\{1,2\}}, M_{\{3,4\}}^{\{1,3\}}, M_{\{3,4\}}^{\{2,3\}}, \\ M_{\{1\}}^{\{1,3\}}, M_{\{1\}}^{\{2,3\}}, M_{\{2\}}^{\{3\}}, M_{\{2\}}^{\{1,3\}}, M_{\{3\}}^{\{2\}}, M_{\{3\}}^{\{2,3\}}, M_{\{4\}}^{\{1\}}, M_{\{4\}}^{\{2\}}, M_{\{4\}}^{\{1,2\}} \end{array} \right\} \tag{35}$$

The network is depicted in Fig. 5. Note that according to the definition (26), the message set $\mathbb{M}$ in (35) is partitioned into the following subsets:

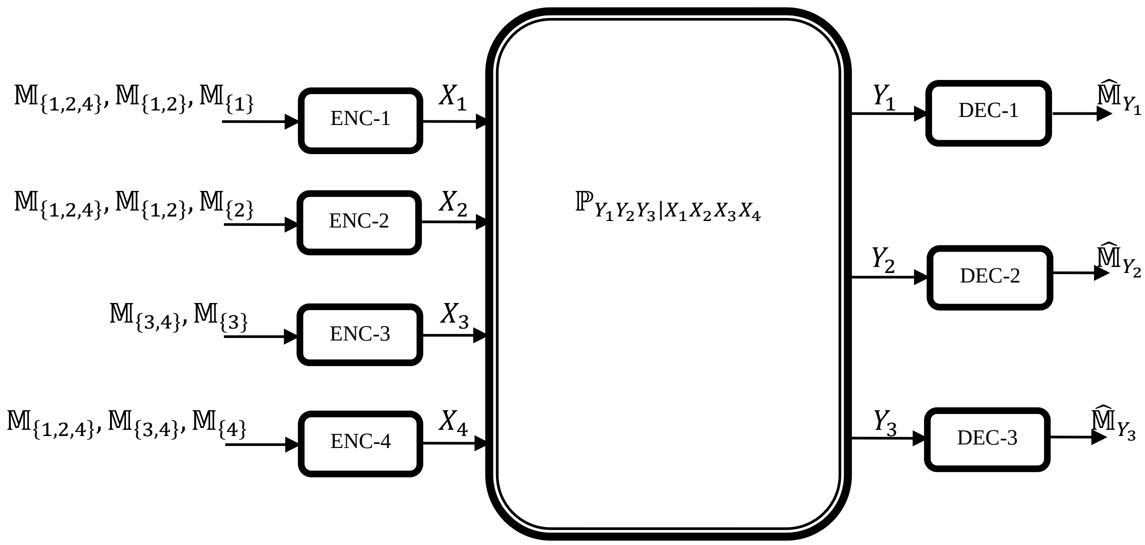

Figure 5. A 4-transmitter/3-receiver interference network with messages in (35).

$$\begin{cases} \mathbb{M}_{\{1,2,4\}} = \left\{M_{\{1,2,4\}}^{\{3\}}, M_{\{1,2,4\}}^{\{1,3\}}, M_{\{1,2,4\}}^{\{2,3\}}\right\} \\ \mathbb{M}_{\{1,2\}} = \left\{M_{\{1,2\}}^{\{2,3\}}, M_{\{1,2\}}^{\{1,3\}}\right\}, \mathbb{M}_{\{3,4\}} = \left\{M_{\{3,4\}}^{\{1,2\}}, M_{\{3,4\}}^{\{1,3\}}, M_{\{3,4\}}^{\{2,3\}}\right\} \\ \mathbb{M}_{\{1\}} = \left\{M_{\{1\}}^{\{1,3\}}, M_{\{1\}}^{\{2,3\}}\right\}, \mathbb{M}_{\{2\}} = \left\{M_{\{2\}}^{\{3\}}, M_{\{2\}}^{\{1,3\}}\right\}, \mathbb{M}_{\{3\}} = \left\{M_{\{3\}}^{\{3\}}, M_{\{3\}}^{\{2,3\}}\right\}, \mathbb{M}_{\{4\}} = \left\{M_{\{4\}}^{\{1\}}, M_{\{4\}}^{\{2\}}, M_{\{4\}}^{\{1,2\}}\right\} \end{cases} \tag{36}$$

As we see these subsets all have more than one element. Nevertheless, in order to achieve the sum-rate capacity it is optimal to transmit only one (carefully picked) message from each set. Using the definitions (27) and (28), we choose the desired messages as follows:

$$\widetilde{\mathbb{M}} = \left\{M_{\{1,2,4\}}^{\{3\}}, M_{\{1,2\}}^{\{2,3\}}, M_{\{3,4\}}^{\{1,2\}}, M_{\{1\}}^{\{1,3\}}, M_{\{2\}}^{\{3\}}, M_{\{3\}}^{\{3\}}, M_{\{4\}}^{\{1\}}\right\} \tag{37}$$

Therefore, according to (28), we have:

$$\begin{cases} \widetilde{\mathbb{M}}_{Y_1} = \left\{M_{\{3,4\}}^{\{1,2\}}, M_{\{1\}}^{\{1,3\}}, M_{\{4\}}^{\{1\}}\right\} \\ \widetilde{\mathbb{M}}_{Y_2} = \left\{M_{\{1,2\}}^{\{2,3\}}, M_{\{3,4\}}^{\{1,2\}}\right\} \\ \widetilde{\mathbb{M}}_{Y_3} = \left\{M_{\{1,2,4\}}^{\{3\}}, M_{\{1,2\}}^{\{2,3\}}, M_{\{1\}}^{\{1,3\}}, M_{\{2\}}^{\{3\}}, M_{\{3\}}^{\{3\}}\right\} \end{cases} \tag{38}$$

The plan of messages corresponding to the set $\widetilde{\mathbb{M}}$ given in (37) is depicted in Fig. 6.

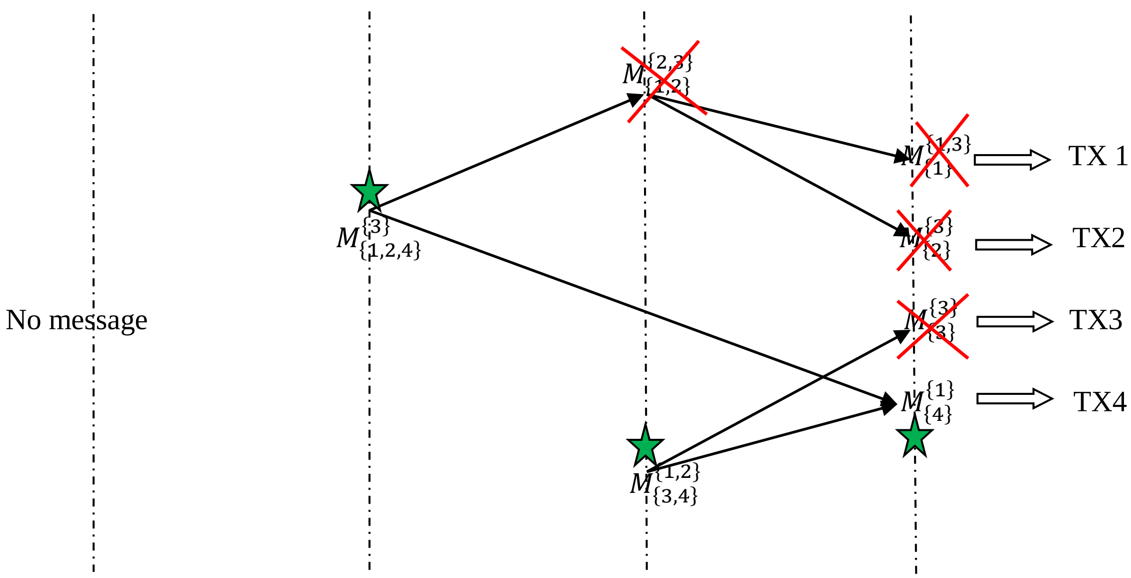

Figure 6. The plan of messages corresponding to the messages (37) for Example 3.1. To achieve the sum-rate capacity of the degraded network, it is only required to transmit the messages denoted by a green star and ignore the others.

Also, based on (29), we have:

$$\begin{cases} \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_1}} = \left\{M_{\{4\}}^{\{1\}}\right\} \\ \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_2}} = \left\{M_{\{3,4\}}^{\{1,2\}}\right\} \\ \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_3}} = \left\{M_{\{1,2,4\}}^{\{3\}}, M_{\{1,2\}}^{\{2,3\}}, M_{\{1\}}^{\{1,3\}}, M_{\{2\}}^{\{3\}}, M_{\{3\}}^{\{3\}}\right\} \end{cases} \tag{39}$$

Now we can determine the sets $\widetilde{\mathbb{M}}_{Y_1}^{*}, \widetilde{\mathbb{M}}_{Y_2}^{*}$, and $\widetilde{\mathbb{M}}_{Y_3}^{*}$ as follows (see definitions (31, 32)). Considering the plan of messages in Fig. 6,

1. The message $M_{\{1,2,4\}}^{\{3\}}$ is a cloud center for the messages $M_{\{1,2\}}^{\{2,3\}}, M_{\{1\}}^{\{1,3\}}, M_{\{2\}}^{\{3\}}$. Also, the message $M_{\{3,4\}}^{\{1,2\}}$ is a cloud center for $M_{\{3\}}^{\{3\}}$. Therefore, from the set $\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_3}}$ only the message $M_{\{1,2,4\}}^{\{3\}}$ belongs to $\widetilde{\mathbb{M}}_{Y_3}^{*}$. Thus, $\widetilde{\mathbb{M}}_{Y_3}^{*} = \left\{M_{\{1,2,4\}}^{\{3\}}\right\}$.
2. There is no cloud center for the message $M_{\{3,4\}}^{\{1,2\}}$ in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_3}}\right) = \left\{M_{\{3,4\}}^{\{1,2\}}, M_{\{4\}}^{\{1\}}\right\}$. Thus, $\widetilde{\mathbb{M}}_{Y_2}^{*} = \left\{M_{\{3,4\}}^{\{1,2\}}\right\}$.
3. The message $M_{\{4\}}^{\{1\}}$ is a satellite for both $M_{\{1,2,4\}}^{\{3\}}$ and $M_{\{3,4\}}^{\{1,2\}}$, but not for a message in $\widetilde{\mathbb{M}} - \left(\widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_2}} \cup \widetilde{\mathbb{M}}_{\overset{\mapsto}{Y_3}}\right) = \left\{M_{\{4\}}^{\{1\}}\right\}$. Thus, $\widetilde{\mathbb{M}}_{Y_1}^{*} = \left\{M_{\{4\}}^{\{1\}}\right\}$.

In Fig. 6, we have marked the messages belonging to $\mathbb{M}^{*} = \widetilde{\mathbb{M}}_{Y_3}^{*} \cup \widetilde{\mathbb{M}}_{Y_2}^{*} \cup \widetilde{\mathbb{M}}_{Y_1}^{*}$ by a green star and those in $\mathbb{M}^{\times} = \mathbb{M}_{Y_1}^{\times} \cup \mathbb{M}_{Y_2}^{\times} \cup \mathbb{M}_{Y_3}^{\times}$ by a red multiplication sign. Based on Theorem 3.2, only those messages which are denoted by green star are effective in computation of the sum-rate capacity. Therefore, the sum-rate capacity of the degraded network is given by:

$$\mathcal{C}_{sum}^{EX\ 3.1} = \max_{\overset{*}{\mathcal{P}}_{sum}^{EX\ 3.1}} \left(I\left(M_{\{4\}}^{\{1\}}; Y_1 \middle| M_{\{3,4\}}^{\{1,2\}}, M_{\{1,2,4\}}^{\{3\}}, Q\right) + I\left(M_{\{3,4\}}^{\{1,2\}}; Y_2 \middle| M_{\{1,2,4\}}^{\{3\}}, Q\right) + I\left(M_{\{1,2,4\}}^{\{3\}}; Y_3 \middle| Q\right)\right) \tag{40}$$

where $\overset{*}{\mathcal{P}}_{sum}^{EX\ 3.1}$ denotes the set of all joint PDFs as:

$$P_Q \times P_{M_{\{1,2,4\}}^{\{3\}}} \times P_{M_{\{3,4\}}^{\{1,2\}}} \times P_{M_{\{4\}}^{\{1\}}} \times P_{X_1 | M_{\{1,2,4\}}^{\{3\}}, Q} \times P_{X_2 | M_{\{1,2,4\}}^{\{3\}}, Q} \times P_{X_3 | M_{\{3,4\}}^{\{1,2\}}, Q} \times P_{X_4 | M_{\{4\}}^{\{1\}}, M_{\{1,2,4\}}^{\{3\}}, M_{\{3,4\}}^{\{1,2\}} Q} \tag{41}$$

For brevity of notations, let denote the auxiliaries $M_{\{3,4\}}^{\{1,2\}}$ and $M_{\{1,2,4\}}^{\{3\}}$ by U and V, respectively. Since X_4 is a deterministic function of $M_{\{4\}}^{\{1\}}, M_{\{1,2,4\}}^{\{3\}}$, and $M_{\{3,4\}}^{\{1,2\}}$, the sum capacity expression (40) can be indeed re-expressed as:

$$\mathcal{C}_{sum}^{EX\ 3.1} = \max_{\mathcal{P}_{sum}^{EX\ 3.1}} \left(I(X_4; Y_1 | U, V, Q) + I(U; Y_2 | V, Q) + I(V; Y_3 | Q)\right) \tag{42}$$

where $\mathcal{P}_{sum}^{EX\,3.1}$ denotes the set of all joint PDFs given below:

$$P_Q P_{U|Q} P_{V|Q} P_{X_1|VQ} P_{X_2|VQ} P_{X_3|UQ} P_{X_4|UVQ}, \qquad P_{X_1|VQ}, P_{X_2|VQ}, P_{X_3|UQ} \in \{0,1\} \tag{43}$$

The expression (42) can be further simplified as follows:

$$\begin{aligned}
\mathcal{C}_{sum}^{EX\,3.1} &= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \left(I(X_4;Y_1|U,V,Q) + I(U;Y_2|V,Q) + I(V;Y_3|Q)\right) \\
&\overset{(a)}{=} \max_{\mathcal{P}_{sum}^{EX\,3.1}} \left(I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(X_3,U;Y_2|X_1,X_2,V,Q) + I(X_1,X_2,V;Y_3|Q)\right) \\
&= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \begin{pmatrix} I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(X_3,U;Y_2|X_1,X_2,V,Q) \\ +I(V;Y_3|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q) \end{pmatrix} \\
&\overset{(b)}{\leq} \max_{\mathcal{P}_{sum}^{EX\,3.1}} \begin{pmatrix} I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(X_3,U;Y_2|X_1,X_2,V,Q) \\ +I(V;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q) \end{pmatrix} \\
&= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \left(I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(X_3,U,V;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q)\right) \\
&= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \begin{pmatrix} I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(U,V;Y_2|X_1,X_2,X_3,Q) \\ +I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q) \end{pmatrix} \\
&\overset{(c)}{\leq} \max_{\mathcal{P}_{sum}^{EX\,3.1}} \begin{pmatrix} I(X_4;Y_1|X_1,X_2,X_3,U,V,Q) + I(U,V;Y_1|X_1,X_2,X_3,Q) \\ +I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q) \end{pmatrix} \\
&= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \left(I(X_4,U,V;Y_1|X_1,X_2,X_3,Q) + I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q)\right) \\
&= \max_{\mathcal{P}_{sum}^{EX\,3.1}} \left(I(X_4;Y_1|X_1,X_2,X_3,Q) + I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q)\right)
\end{aligned} \tag{44}$$

where equality (a) holds because X_1 and X_2 are given by deterministic functions of (V,Q), and X_3 is given by a deterministic function of (U,Q); also the inequalities (b) and (c) are due to the degradedness of the network. Note that according to (43), conditioning on Q, the pair (X_1,X_2) is independent of X_3. On the other hand, it is clear that by setting $U \equiv X_3$ and $V \equiv (X_1,X_2)$ in (42), the expression at the right side of the last equality in (44) is derived. Thus, the sum-rate capacity is given by:

$$\mathcal{C}_{sum}^{EX\,3.1} = \max_{P_Q P_{X_1X_2|Q} P_{X_3|Q} P_{X_4|X_1X_2X_3Q}} \left(I(X_4;Y_1|X_1,X_2,X_3,Q) + I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q)\right) \tag{45}$$

Now let us consider the Gaussian version of Example 3.1. The degraded network is formulated as follows:

$$\begin{cases} Y_1 = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + Z_1 \\ Y_2 = \dfrac{a_1}{b_2}X_1 + \dfrac{a_2}{b_2}X_2 + \dfrac{a_3}{b_2}X_3 + \dfrac{a_4}{b_2}X_4 + Z_2 \qquad , \qquad b_2, b_3 \geq 1 \\ Y_3 = \dfrac{a_1}{b_2b_3}X_1 + \dfrac{a_2}{b_2b_3}X_2 + \dfrac{a_3}{b_2b_3}X_3 + \dfrac{a_4}{b_2b_3}X_4 + Z_3 \end{cases} \tag{46}$$

where Z_1, Z_2, Z_3 are Gaussian noises with zero mean and unit variance, and the input signals are subject to power constraints $\mathbb{E}[X_i^2] \leq P_i, i = 1,2,3,4$. The channel (46) is in fact equivalent to the following:

$$\begin{cases} \tilde{Y}_1 = Y_1 \\ \tilde{Y}_2 = \frac{1}{b_2}\tilde{Y}_1 + \sqrt{1 - \frac{1}{b_2^2}}\,\tilde{Z}_2 \\ \tilde{Y}_3 = \frac{1}{b_3}\tilde{Y}_2 + \sqrt{1 - \frac{1}{b_3^2}}\,\tilde{Z}_3 \end{cases} \tag{47}$$

where $\tilde{Z}_2, \tilde{Z}_3$ are independent Gaussian RVs (also independent of Z_1) with zero mean and unit variance. We now present an explicit characterization for the sum-rate capacity of the Gaussian network in (46).

Proposition 3.1) *Consider the multi-message interference network in Fig. 5. The sum-rate capacity of the Gaussian network in (46) is given in the following:*

$$\mathcal{C}_{sum}^{EX.1\sim G} = \max_{\substack{\alpha,\beta\in[0,1] \\ \alpha^2+\beta^2\leq 1}} \left(\begin{array}{c} \psi\left(a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + \psi\left(\dfrac{\frac{1}{b_2^2}\left(|a_3| + |a_4|\beta\sqrt{\frac{P_4}{P_3}}\right)^2 P_3}{\frac{a_4^2}{b_2^2}\left(1-(\alpha^2+\beta^2)\right)P_4 + 1} \right) \\ +\psi\left(\dfrac{\frac{1}{b_2^2b_3^2}\left(a_1^2P_1 + a_2^2P_2 + a_4^2\alpha^2P_4 + 2|a_1a_2|\sqrt{P_1P_2} + 2|a_1a_4|\alpha\sqrt{P_1P_4} + 2|a_2a_4|\alpha\sqrt{P_2P_4}\right)}{\frac{1}{b_2^2b_3^2}\left(\left(|a_3| + |a_4|\beta\sqrt{\frac{P_4}{P_3}}\right)^2 P_3 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1} \right) \end{array} \right) \tag{48}$$

Proof of Proposition 3.1) Let X_1, X_3 and Z be independent Gaussian RVs with zero means and variances P_1, P_3, and 1, respectively. Moreover, define X_2 and X_4 as follows:

$$\begin{cases} X_2 \triangleq \mathrm{sign}(a_1 a_2)\sqrt{\frac{P_2}{P_1}}X_1 \\ X_4 \triangleq \mathrm{sign}(a_1 a_4)\alpha\sqrt{\frac{P_4}{P_1}}X_1 + \mathrm{sign}(a_3 a_4)\beta\sqrt{\frac{P_4}{P_3}}X_3 + \sqrt{\left(1-(\alpha^2+\beta^2)\right)P_4}Z \end{cases} \tag{49}$$

where $\alpha, \beta \in [0,1]$ are arbitrary real numbers with $\alpha^2 + \beta^2 \leq 1$; also, for a real number x, $\mathrm{sign}(x)$ is equal to 1 if x is nonnegative and -1 otherwise. Then, by setting X_1, X_2, X_3, and X_4 as given by (49) and $Q \equiv \emptyset$ in (45), we derive the achievability of (48). The proof of the optimality of Gaussian input distributions is given in the Appendix A. ■

Remark: The method of our proof presented in Appendix A for the optimality of Gaussian distributions to achieve the sum-rate capacity of the Gaussian network of Example 3.1 can be adapted to other scenarios. Nonetheless, as the network is degraded, by manipulating the mutual information functions in (33), one may present other arguments for the optimality of Gaussian inputs.

Example 3.2: K-User Classical Interference Channel

Consider the K-User classical IC as shown in Fig. 7. In this network, each transmitter sends a message to its respective receiver and there is no cooperation among transmitters.

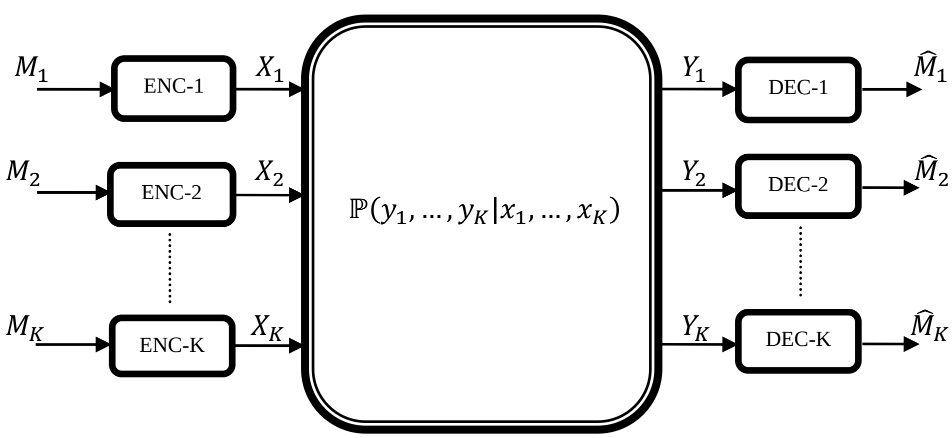

Figure 7. The K-user classical interferecne channel.

The sum-rate capacity of the degraded network is directly derived from Theorem 3.1, which is given as follows:

$$\mathcal{C}_{sum}^{IC_{deg}} = \max_{\mathcal{P}_{sum}^{IC_{deg}}} \big(I(M_1;Y_1|M_2,\dots,M_{K-1},M_K,Q)+\cdots+I(M_{K-1};Y_{K-1}|M_K,Q)+I(M_K;Y_K|Q)\big) \tag{50}$$

where $\mathcal{P}_{sum}^{IC_{deg}}$ is the set of all joint PDFs given by:

$$P_Q\prod_{l=1}^{K} P_{M_l}P_{X_l|M_l,Q} \tag{51}$$

Also, $P_{X_l|M_l,Q}\in\{0,1\}, l=1,\dots,K$; in other words, X_l is a deterministic function of (M_l,Q). In fact, one can readily re-derive the sum-rate capacity as follows:

$$\mathcal{C}_{sum}^{IC_{deg}} = \max_{P_Q\prod_{l=1}^{K}P_{X_l|Q}} \big(I(X_1;Y_1|X_2,\dots,X_{K-1},X_K,Q)+\cdots+I(X_{K-1};Y_{K-1}|X_K,Q)+I(X_K;Y_K|Q)\big) \tag{52}$$

It is remarkable that this is the first sum-rate capacity result for the K-user classical IC which is derived for *both discrete and Gaussian networks*. The sum-rate capacity of the degraded Gaussian IC was derived in [5] using a rather complicated approach (based on genie-aided techniques). It is not difficult to show that for the Gaussian network, the optimal solution to the optimization (52) is attained by Gaussian distributions.

Example 3.3: As we see from the expressions (45) and (52) for the scenarios of Examples 3.1 and 3.2, to describe the sum-rate capacity there is no need to make use of auxiliary random variables. One may think that the same consequence holds for other networks. In fact, this is the case for many network scenarios but not for the general case. Here, we provide an example where it is essential to use auxiliary random variables. Specifically, consider the 2-transmitter/2-receiver interference network shown in Fig. 8. In this network, two transmitters cooperatively send a message to the second receiver, and also each transmitter sends separately a message to the first receiver. Using (32), one can readily show that for this network $\mathbb{M}^*=\mathbb{M}$. The plan of messages for this scenario is depicted in Fig. 9.

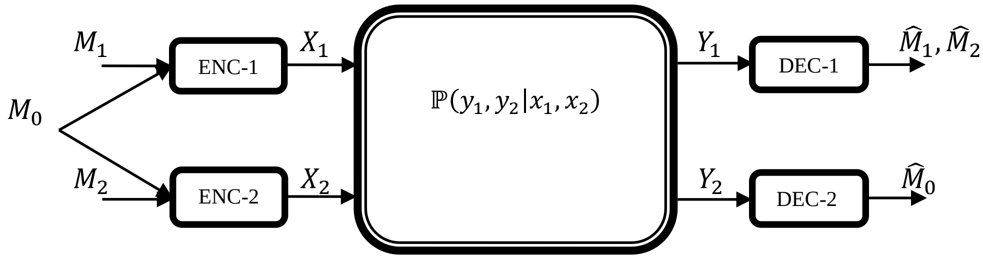

Figure 8. A 2-transmitter/2-receiver interference network.

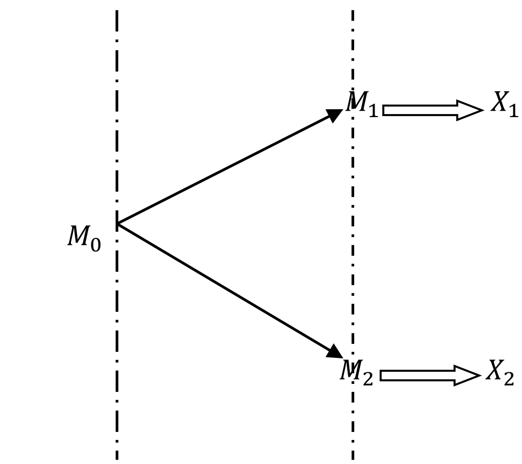

Figure 9. The plan of messages for the interference network of Fig. 8.

Based on Theorem 3.1 and 3.2, the sum-rate capacity of the degraded channel is given by:

$$\mathcal{C}_{sum}^{EX\ 3.3} = \max_{\mathcal{P}_{sum}^{EX\ 3.3}} \big(I(M_1, M_2; Y_1|M_0, Q) + I(M_0; Y_2|Q)\big) \tag{53}$$

where $\mathcal{P}_{sum}^{EX\ 3.3}$ denotes the set of all joint PDFs of the form below:

$$P_Q P_{M_0} P_{M_1} P_{M_2} P_{X_1|M_0 M_1 Q} P_{X_2|M_0 M_2 Q} \tag{54}$$

One can easily show that the sum-rate capacity can be re-written as follows:

$$\mathcal{C}_{sum}^{EX\ 3.3} = \max_{P_Q P_{W|Q} P_{X_1|WQ} P_{X_2|WQ}} \left(I(X_1, X_2; Y_1|W, Q) + I(W; Y_2|Q)\right)$$
$$= \max_{P_W P_{X_1|W} P_{X_2|W}} \left(I(X_1, X_2; Y_1|W) + I(W; Y_2)\right) \tag{55}$$

In general, none of the choices $W \equiv X_1$, $W \equiv X_2$ or $W \equiv \emptyset$ is optimal for the right side of (55). Therefore, it is unavoidable to use the auxiliary random variable W to describe the sum-rate capacity. Let us now consider the Gaussian channel which is formulated as follows:

$$\begin{cases} Y_1 = X_1 + aX_2 + Z_1 \\ Y_2 = bX_1 + X_2 + Z_2 \end{cases} \tag{56}$$

where Z_1 and Z_2 are independent Gaussian random variables with zero means and unit variances. The channel is degraded (Y_2 is a degraded version of Y_1) provided that $ab = 1$ as well as $|a| \geq 1$. In this case, the channel is equivalent to the following:

$$\begin{cases} Y_1 = X_1 + aX_2 + Z_1 \\ \tilde{Y}_2 = bY_1 + \sqrt{1 - b^2}\, \tilde{Z}_2 \end{cases} \tag{57}$$

where $\tilde{Z}_2$ is Gaussian RV with zero mean and unit variance and independent of Z_1. In the next proposition, we derive the sum-rate capacity for this network.

Proposition 3.2) *The sum-rate capacity of the Gaussian interference network (56) shown in Fig. 8 with* $ab = 1$ *and* $|a| \geq 1$ *is given below:*

$$\mathcal{C}_{sum}^{EX\ 3.3\sim G} = \max_{-1 \leq \alpha, \beta \leq 1} \begin{pmatrix} \psi\left((1 - \alpha^2)P_1 + a^2(1 - \beta^2)P_2\right) \\ +\psi\left(\dfrac{b^2\alpha^2 P_1 + \beta^2 P_2 + 2b\alpha\beta\sqrt{P_1 P_2}}{b^2(1 - \alpha^2)P_1 + (1 - \beta^2)P_2 + 1}\right) \end{pmatrix} \tag{58}$$

Proof of Proposition 3.2) Let $W, \tilde{X}_1, \tilde{X}_2$ be independent Gaussian RVs with zero mean and unit variances. Define:

$$\begin{cases} X_1 \triangleq \alpha\sqrt{P_1}W + \sqrt{(1 - \alpha^2)P_1}\tilde{X}_1 \\ X_2 \triangleq \beta\sqrt{P_2}W + \sqrt{(1 - \beta^2)P_2}\tilde{X}_2 \end{cases}, \qquad -1 \leq \alpha, \beta \leq 1 \tag{59}$$

By setting W, X_1 and X_2 in (55), we obtain the achievability. The converse proof is given in Appendix B. ■

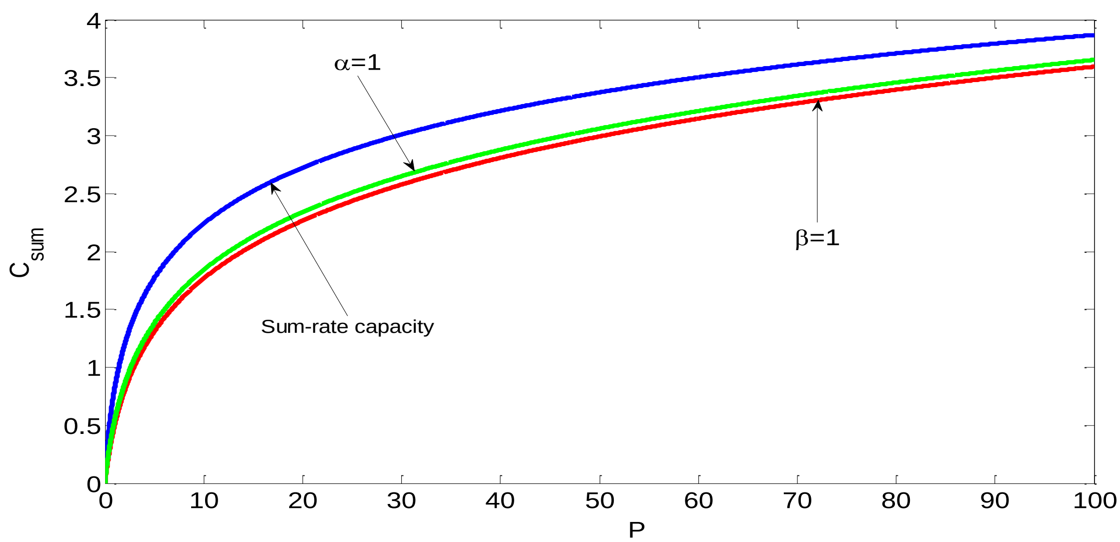

Figure 10. The sum-rate capcaity of the Gaussian degraded interference network (56) of Fig. 8 in terms of P, where $P = P_1 = 200P_2$, and $a = {}^{1}/_{b} = 15$.

In Fig. 10, we have numerically evaluated the sum-rate capacity for the Gaussian degraded interference network (56) of Fig. 8 in terms of P, where $P = P_1 = 200P_2$. Also, we have depicted the expression in the right side of (58) for $\alpha = 1$ as well as for $\beta = 1$. As we see, both choices are sub-optimal. Therefore, both $W \equiv X_1$ and $W \equiv X_2$ in (55) are suboptimal, verifying that the use of auxiliary random variable is unavoidable to describe the sum-rate capacity.

Chapter 4

General Interference Networks: Unified Outer Bounds

One of the origins of the difficulty in the analysis of the interference networks, specifically networks with more than two receivers, is the lack of non-trivial capacity outer bounds with a satisfactory performance. In this chapter, we directly make use of the general expression derived in Theorem 3.2 for the sum-rate capacity of degraded networks to establish a unified outer bound on the sum-rate capacity of the general non-degraded interference networks in Fig. 2. We then present several classes of non-degraded networks for which the derived outer bound is sum-rate optimal.

A main result of this chapter is given in the following theorem.

Theorem 4.1: *Consider the K_1-Transmitter/K_2-Receiver interference network in Fig. 2 with the message sets $\mathbb{M}, \mathbb{M}_{X_i}, i = 1, \dots, K_1$, and $\mathbb{M}_{Y_j}, j = 1, \dots, K_2$. The sum-rate capacity is outer-bounded by:*

$$\mathcal{C}_{sum}^{GIN} \leq \max_{\overset{*}{\mathcal{P}}_{sum}^{GIN}} \Bigg(I\left(\mathbb{M}_{Y_1}^*; \bar{\bar{Y}}_1 \middle| \mathbb{M}_{Y_2}^*, \dots, \mathbb{M}_{Y_{K_2-1}}^*, \mathbb{M}_{Y_{K_2}}^*, Q\right) + \cdots + I\left(\mathbb{M}_{Y_{K_2-1}}^*; \bar{\bar{Y}}_{K_2-1} \middle| \mathbb{M}_{Y_{K_2}}^*, Q\right) + I\left(\mathbb{M}_{Y_{K_2}}^*; \bar{\bar{Y}}_{K_2} \middle| Q\right) \Bigg) \tag{60}$$

where $\bar{\bar{Y}}_1, \bar{\bar{Y}}_2, \dots, \bar{\bar{Y}}_{K_2}$ are given below:

$$\bar{\bar{Y}}_j \triangleq \left(Y_j, Y_{j+1}, \dots, Y_{K_2}\right), \qquad j = 1, \dots, K_2 \tag{61}$$

also, $\overset{*}{\mathcal{P}}_{sum}^{GIN}$ denotes the set of all joint PDFs as:

$$P_Q \times \prod_{M_\Delta^\nabla \in \mathbb{M}^*} P_{M_\Delta^\nabla} \times \prod_{i \in [1:K_1]} P_{X_i | \mathbb{M}_{X_i}^* Q} \tag{62}$$

Moreover, the PDFs $P_{M_\Delta^\nabla}$, $M_\Delta^\nabla \in \mathbb{M}^*$ are uniformly distributed, and $P_{X_i | \mathbb{M}_{X_i}^* Q} \in \{0,1\}$ for $i = 1, \dots, K_1$, i.e., X_i is a deterministic function of $\left(\mathbb{M}_{X_i}^*, Q\right)$. The messages $\mathbb{M}^*$ are given in (31) and (32); in other words, the messages are determined based on the procedure we developed to achieve the sum-rate capacity of the degraded networks in Chapter 3.

Proof of Theorem 4.1) The proof is easily understood from the structure of the outer bound as well as the sum-rate capacity of the degraded networks given in Theorem 3.2. Consider an artificial network with receivers $\bar{\bar{Y}}_1, \bar{\bar{Y}}_2, \dots, \bar{\bar{Y}}_{K_2}$ as given in (61). It is clear that the capacity region of this artificial network contains that of the original network as a subset. Moreover, the artificial network is degraded in the sense of (12). Therefore, according to Theorem 3.2, its sum-rate capacity is given by (60). The proof is thus complete. ■

Remark: Let $\lambda(.)$ be an arbitrary permutation of the elements of the set $\{1, \dots, K_2\}$. By exchanging the indices $1, \dots, K_2$ with $\lambda(1), \dots, \lambda(K_2)$ in (60), respectively, we indeed derive other outer bounds on the sum-rate capacity. Although, it is very important to note that the messages $\mathbb{M}^*$ and the corresponding subsets $\mathbb{M}^*_{X_i}, i = 1, \dots, K_1$, and $\mathbb{M}^*_{Y_j}, j = 1, \dots, K_2$, may vary by the permutation $\lambda(.)$; for each permutation $\lambda(.)$, the messages $\mathbb{M}^*$ should be discriminated from the set $\mathbb{M}$ based on the order of the degradedness of the corresponding artificial network.

Let us provide an example. Consider the network of Example 3.1 shown in Fig. 5 but without the condition of being degraded. Based on Theorem 4.1, the following constitutes an outer bound on the sum-rate capacity:

$$\mathcal{C}_{sum}^{IN\to\text{Fig. 5}} \leq$$

$$\max_{P_Q P_{X_1 X_2|Q} P_{X_3|Q} P_{X_4|X_1 X_2 X_3 Q}} \left(I(X_4; Y_1, Y_2, Y_3|X_1, X_2, X_3, Q) + I(X_3; Y_2, Y_3|X_1, X_2, Q) + I(X_1, X_2; Y_3|Q) \right) \tag{63}$$

It is clear that the outer bound (63) is tighter than the standard cut-set outer bound. In fact, five other bounds similar to (63) can be established on the sum-rate capacity.

In the following subchapter, we show that the outer bound (60) is indeed sum-rate optimal for many different interference networks.

4.1 Multiple-Access-Interference Networks (MAIN)

As indicated by Example 3.3, to describe the bound (60) for the general interference network, the use of auxiliary random variables is unavoidable. Nevertheless, in what follows, we will present classes of networks for which one can derive a representation of the outer bound (60) without any auxiliary random variable.

Consider network scenarios where some groups of transmitters interested in transmitting information to their respective receiver while causing interference to the other receivers. Clearly, consider an interference network with $K_1 = \sum_{l=1}^{K_2} \mu_l$ transmitters and K_2 receivers, where $\mu_1, \dots, \mu_{K_2}$ are arbitrary natural numbers. The transmitters are partitioned into K_2 sets labeled $\mathbb{X}_1, \dots, \mathbb{X}_{K_2}$ such that those in $\mathbb{X}_j = \left\{X_{j,1}, \dots, X_{j,\mu_j}\right\}$ send the messages $\mathbb{M}_{Y_j}$ to the receiver $Y_j, j = 1, \dots, K_2$, but they have no message for the other receivers; in other words, the message sets $\mathbb{M}_{Y_1}, \dots, \mathbb{M}_{Y_{K_2}}$ are pairwise disjoint. Also, the distribution of messages $\mathbb{M}_{Y_j}$ among the transmitters $\mathbb{X}_j = \left\{X_{j,1}, \dots, X_{j,\mu_j}\right\}$ is arbitrary. The network model has been shown in Fig. 11.

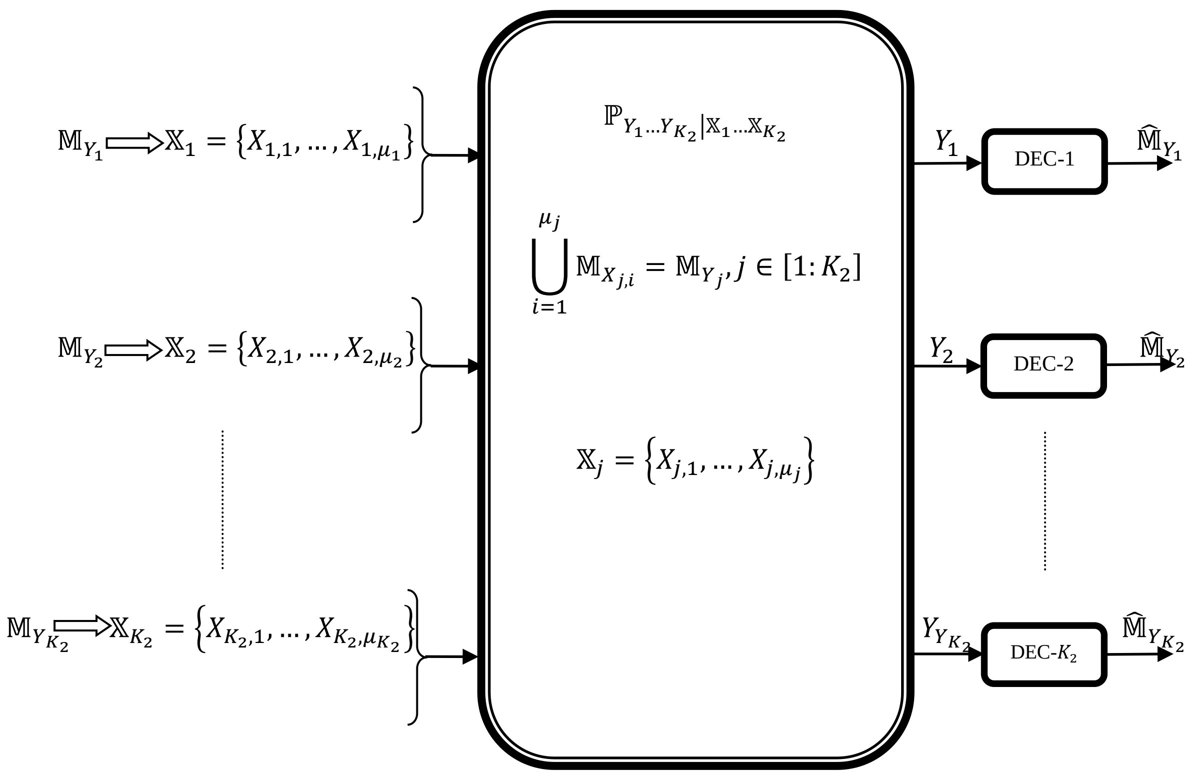

Figure 11. The general Multiple-Access-Interference Networks (MAIN). For $j = 1, \dots, K_2$, $\mathbb{X}_j$ denotes a set of arbitrary transmitters which send (in an arbitrary order) the messages $\mathbb{M}_{Y_j}$ to the receiver Y_j.

These scenarios do not contain broadcasting messages to multiple receivers because each transmitter sends information only to a single receiver. In fact, such networks are composed of several interfering multiple access channels; hence, we call them as *Multiple-Access-Interference Networks (MAIN)*. In the following proposition, we prove that for the MAINs, the outer bound (60) can be simply represented by using only the input and output signals.

Proposition 4.1) *Consider the general MAIN shown in Fig. 11. For this network the sum-rate outer bound in (60) is simplified as follows:*

$$\mathcal{C}_{sum}^{MAIN} \leq \max_{\overset{*}{\mathcal{P}}_{sum}^{MAIN}} \left(I\left(\mathbb{X}_1; \bar{\bar{Y}}_1 \middle| \mathbb{X}_2, \dots, \mathbb{X}_{K_2-1}, \mathbb{X}_{K_2}, Q\right) + \cdots + I\left(\mathbb{X}_{K_2-1}; \bar{\bar{Y}}_{K_2-1} \middle| \mathbb{X}_{K_2}, Q\right) + I\left(\mathbb{X}_{K_2}; \bar{\bar{Y}}_{K_2} \middle| Q\right) \right) \tag{64}$$

where $\bar{\bar{Y}}_1, \bar{\bar{Y}}_2, \dots, \bar{\bar{Y}}_{K_2}$ are given in (61), and $\overset{*}{\mathcal{P}}_{sum}^{MAIN}$ denotes the set of all joint PDFs which are induced on the input signals $Q, \mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_{K_2-1}, \mathbb{X}_{K_2}$, by the following PDFs:

$$P_Q \times \prod_{M_\Delta^\nabla \in \mathbb{M}^*} P_{M_\Delta^\nabla} \times \prod_{i \in \left[1 : \sum_{l=1}^{K_2} \mu_l\right]} P_{X_i | \mathbb{M}_{X_i}^* Q} \tag{65}$$

The messages $\mathbb{M}^*$ are a subset of $\mathbb{M}$ which are determined using the algorithm we presented in Chapter 3. In fact, considering the plan of messages, $\mathbb{M}^*$ consists of those elements of $\mathbb{M}$ which are not a satellite for any other message. Moreover, if the network transition probability function implies the following Markov chains:

$$\mathbb{X}_j \to Y_j, \mathbb{X}_{j+1}, \dots, \mathbb{X}_{K_2} \to Y_{j+1}, Y_{j+2}, \dots, Y_{K_2}, \qquad j = 1, \dots, K_2 - 1 \tag{66}$$

then, the outer bound is further simplified as:

$$\mathcal{C}_{sum}^{MAIN} \leq \max_{\overset{*}{\mathcal{P}}_{sum}^{MAIN}} \left(I\left(\mathbb{X}_1; Y_1 \middle| \mathbb{X}_2, \dots, \mathbb{X}_{K_2-1}, \mathbb{X}_{K_2}, Q\right) + \cdots + I\left(\mathbb{X}_{K_2-1}; Y_{K_2-1} \middle| \mathbb{X}_{K_2}, Q\right) + I\left(\mathbb{X}_{K_2}; Y_{K_2} \middle| Q\right) \right) \tag{67}$$

Proof of Proposition 4.1) Consider the outer bound (60). For the MAIN scenario in Fig. 11, since the message sets $\mathbb{M}_{Y_1}, \dots, \mathbb{M}_{Y_{K_2}}$ are pairwise disjoint and also each of these sets are sent only to a single receiver, the messages $\mathbb{M}^*$ are those elements of $\mathbb{M}$ which are not a satellite for any other message (considering the plan of messages). Moreover, we have the following joint PDF for the inputs and outputs:

$$P_Q \prod_{M_\Delta^\nabla \in \mathbb{M}^*} P_{M_\Delta^\nabla} \prod_{j=1}^{K_2} P_{\mathbb{X}_j|\mathbb{M}_{Y_j}^* Q} \, \mathbb{P}_{Y_1 \dots Y_{K_2}|\mathbb{X}_1 \dots \mathbb{X}_{K_2}} \tag{68}$$

where $P_{\mathbb{X}_j|\mathbb{M}_{Y_j}^* Q} \in \{0,1\}$. One can readily check that the distribution in (68) implies the following Markov relations:

$$\mathbb{M}_{Y_j}^* \to \mathbb{X}_j, Q \to Y_1, \dots, Y_{K_2}, \qquad j = 1, \dots, K_2 \tag{69}$$

Therefore, for the argument of the maximization in (60), we have:

$$\begin{aligned}
&I\left(\mathbb{M}_{Y_1}^*; \bar{\bar{Y}}_1 \middle| \mathbb{M}_{Y_2}^*, \dots, \mathbb{M}_{Y_{K_2-1}}^*, \mathbb{M}_{Y_{K_2}}^*, Q\right) + \dots + I\left(\mathbb{M}_{Y_{K_2-1}}^*; \bar{\bar{Y}}_{K_2-1} \middle| \mathbb{M}_{Y_{K_2}}^*, Q\right) + I\left(\mathbb{M}_{Y_{K_2}}^*; \bar{\bar{Y}}_{K_2} \middle| Q\right) \\
&\overset{(a)}{=} \left(\begin{array}{c} I\left(\mathbb{X}_1, \mathbb{M}_{Y_1}^*; \bar{\bar{Y}}_1 \middle| \mathbb{M}_{Y_2}^*, \dots, \mathbb{M}_{Y_{K_2}}^*, \mathbb{X}_2, \dots, \mathbb{X}_{K_2}, Q\right) + I\left(\mathbb{X}_2, \mathbb{M}_{Y_2}^*; \bar{\bar{Y}}_2 \middle| \mathbb{M}_{Y_3}^*, \dots, \mathbb{M}_{Y_{K_2}}^*, \mathbb{X}_3, \dots, \mathbb{X}_{K_2}, Q\right) \\ + \dots + I\left(\mathbb{X}_{K_2-1}, \mathbb{M}_{Y_{K_2-1}}^*; \bar{\bar{Y}}_{K_2-1} \middle| \mathbb{M}_{Y_{K_2}}^*, \mathbb{X}_{K_2}, Q\right) + I\left(\mathbb{X}_{K_2}, \mathbb{M}_{Y_{K_2}}^*; \bar{\bar{Y}}_{K_2} \middle| Q\right) \end{array} \right) \\
&\overset{(b)}{=} I\left(\mathbb{X}_1; \bar{\bar{Y}}_1 \middle| \mathbb{X}_2, \dots, \mathbb{X}_{K_2-1}, \mathbb{X}_{K_2}, Q\right) + \dots + I\left(\mathbb{X}_{K_2-1}; \bar{\bar{Y}}_{K_2-1} \middle| \mathbb{X}_{K_2}, Q\right) + I\left(\mathbb{X}_{K_2}; \bar{\bar{Y}}_{K_2} \middle| Q\right)
\end{aligned} \tag{70}$$

where equality (a) holds because the inputs $\mathbb{X}_j$ are given by deterministic functions of $\left(\mathbb{M}_{Y_j}^*, Q\right)$, and the equality (b) is due to (68)-(69). Also, one can readily check that if the Markov relations (66) hold, then each mutual information function in (64) is reduced to the corresponding one in (67). ■

Remarks:

1. Note that no auxiliary random variable is given in the outer bound (64) for the MAIN of Fig. 11. Nevertheless, the joint PDF that is imposed on the input signals should be carefully determined based on the arrangement of the messages $\mathbb{M}^*$ among the transmitters. The plan of messages is very helpful for this propose.
2. It is clear that for the MAINs which satisfy the degraded condition (12), the sum-rate capacity is given by (67).

We now intend to provide several classes of networks, other than the degraded ones, for which the outer bound of Theorem 4.1 is sum-rate optimal. Specifically, we will introduce two new interference networks coined as "*Generalized Z-Interference Networks*" and "*Many-to-One Interference Networks*". We identify noisy interference regimes for these networks. Also, for the first time, interesting networks are introduced for which a combination of successive decoding and treating interference as noise is sum-rate optimal. The results are given in the following subchapters.

4.1.1 Generalized Z-Interference Networks

Consider a special case of the MAINs in Fig. 11 for which the transition probability function of the network is factorized as follows:

$$\mathbb{P}_{Y_1 \ldots Y_{K_2}|\mathbb{X}_1 \ldots \mathbb{X}_{K_2}} = \mathbb{P}_{Y_1|\mathbb{X}_1}\mathbb{P}_{Y_2|\mathbb{X}_1\mathbb{X}_2}\mathbb{P}_{Y_3|\mathbb{X}_1\mathbb{X}_2\mathbb{X}_3} \ldots \mathbb{P}_{Y_{K_2}|\mathbb{X}_1\mathbb{X}_2\mathbb{X}_3 \ldots \mathbb{X}_{K_2}} \tag{71}$$

The network model is shown in Fig. 12. In this figure, each receiver has been linked to its connected transmitters (see Chapter 2.2) by a dashed line.

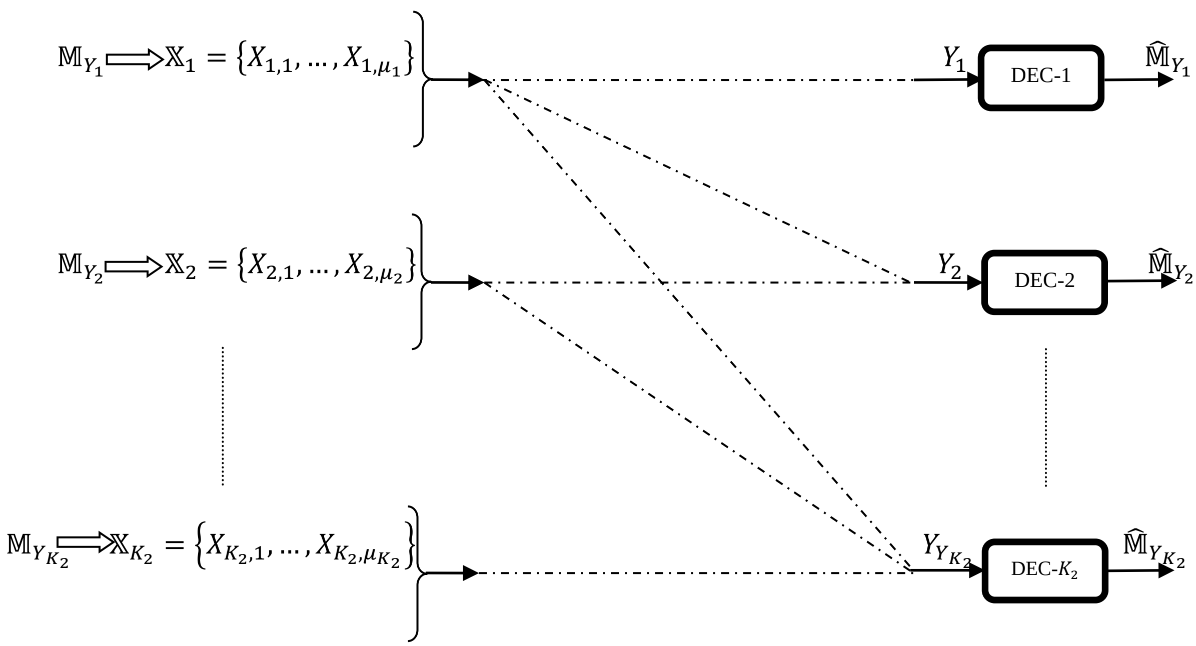

Figure 12. The Generalized Z-Interference Network. For $j = 1, \ldots, K_2$, $\mathbb{X}_j$ denotes a set of arbitrary transmitters which send (in an arbitrary order) the messages $\mathbb{M}_{Y_j}$ to the receiver Y_j.

Note that such networks can be considered as a natural generalization of the Z-interference channel, i.e., the one-sided IC; hence, the name of “Generalized Z-Interference Network”. Now we derive a sum-rate capacity result for these networks. Let the transition probability function of the network satisfies the following degradedness condition as well:

$$\mathbb{P}_{Y_1 \dots Y_{K_2}|\mathbb{X}_1 \dots \mathbb{X}_{K_2}} = \mathbb{P}_{Y_1|\mathbb{X}_1}\mathbb{P}_{Y_2|Y_1\mathbb{X}_2}\mathbb{P}_{Y_3|Y_2\mathbb{X}_3} \cdots \mathbb{P}_{Y_{K_2}|Y_{K_2-1}\mathbb{X}_{K_2}} \tag{72}$$

Note that the degradedness condition in (72) does not imply (12), and the networks in (71)-(72) in general differ from those in (12). In the next theorem, we will prove that for the generalized Z-interference networks given by (72), treating interference as noise achieves the sum-rate capacity.

Theorem 4.2) *Consider the generalized Z-interference network in (71) with the degraded condition (72). The sum-rate capacity of the network is given by the following:*

$$\max_{\overset{*}{\mathcal{P}}_{sum}^{MAIN}} \left(I(\mathbb{X}_1; Y_1|Q) + I(\mathbb{X}_2; Y_2|Q) + \cdots + I\left(\mathbb{X}_{K_2-1}; Y_{K_2-1}\middle|Q\right) + I\left(\mathbb{X}_{K_2}; Y_{K_2}\middle|Q\right) \right) \tag{73}$$

where $\overset{*}{\mathcal{P}}_{sum}^{MAIN}$ is given as in Proposition 4.1.

Proof of Theorem 4.2) The achievability is derived using a simple treating interference as noise strategy: The messages $\mathbb{M} - \mathbb{M}^*$ are withdrawn from the transmission scheme where $\mathbb{M}^*$ are those elements of $\mathbb{M}$ which are not a satellite for any other message (considering the plan of messages). The messages $\mathbb{M}^*$ are encoded separately using independent codewords. The receiver $Y_j, j = 1, \dots, K_2$, jointly decodes its respective messages $\mathbb{M}_{Y_j}^*$ and treats all the other signals as noise; therefore, the following rate is achievable:

$$\max_{\{\text{PDFs in } (65)\}} \left(I\left(\mathbb{M}_{Y_1}^*; Y_1\middle|Q\right) + \cdots + I\left(\mathbb{M}_{Y_{K_2-1}}^*; Y_{K_2-1}\middle|Q\right) + I\left(\mathbb{M}_{Y_{K_2}}^*; Y_{K_2}\middle|Q\right) \right) \tag{74}$$

Similar to the procedure in (70), one can simply show that (74) is equivalent to (73). For the converse part, note that the degradedness condition in (72) implies the Markov relations (66). Therefore, (67) constitutes an outer bound on the sum-rate capacity of the network. Furthermore, for the generalized

Z-interference network (71) shown in Fig. 12, the expression (67) is readily reduced to (73). The proof is thus complete. ■

Remark: The result of Theorem 4.2 indeed holds for all generalized Z-interference networks (71) where the marginal distributions of their transition probability function are given similar to the networks (72).

4.1.2 Many-to-One Interference Networks

We now discuss a special case of the generalized Z-interference networks in (71). Specifically, consider the case where the transition probability function is factorized as follows:

$$\mathbb{P}_{Y_1\ldots Y_{K_2}|\mathbb{X}_1\ldots\mathbb{X}_{K_2}} = \mathbb{P}_{Y_1|\mathbb{X}_1}\mathbb{P}_{Y_2|\mathbb{X}_2}\mathbb{P}_{Y_3|\mathbb{X}_3}\ldots\mathbb{P}_{Y_{K_2-1}|\mathbb{X}_{K_2-1}}\mathbb{P}_{Y_{K_2}|\mathbb{X}_1\mathbb{X}_2\mathbb{X}_3\ldots\mathbb{X}_{K_2}} \tag{75}$$

In other words, only one receiver experiences interference. Fig. 13 depicts the network model where each receiver has been linked to its connected transmitters by a dashed line.

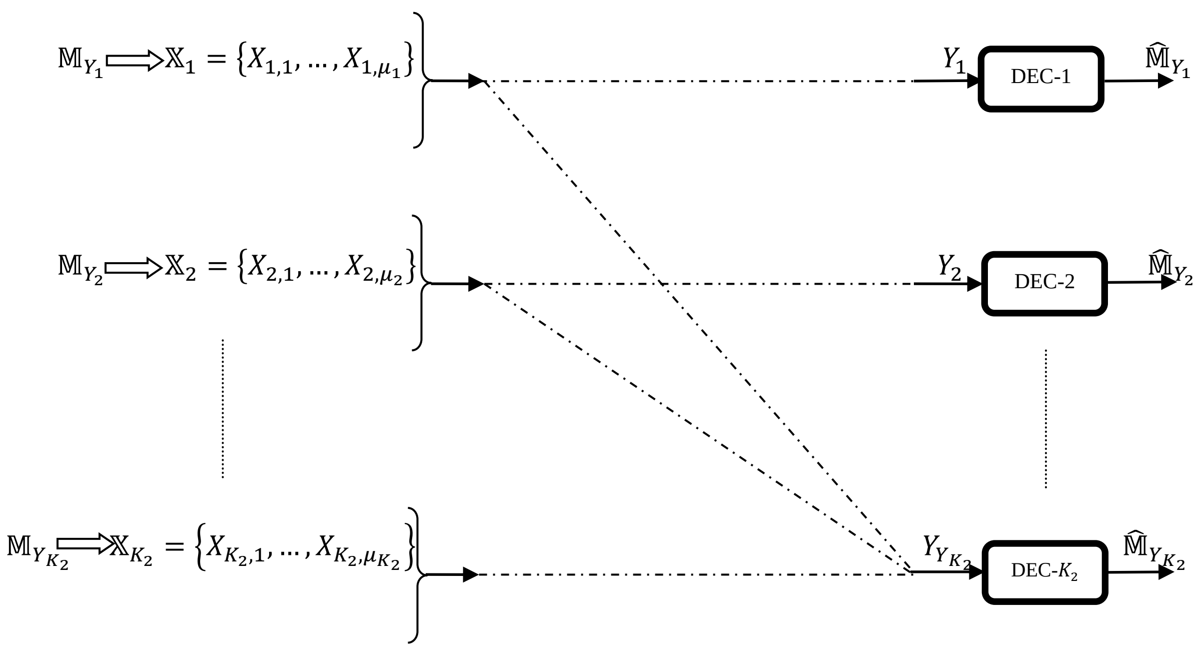

Figure 13. The Many-to-One Interference Network. For $i = 1,\ldots,K_2$, $\mathbb{X}_i$ denotes a set of arbitrary transmitters which send (in an arbitrary order) the messages $\mathbb{M}_{Y_j}$ to the receiver Y_j.

Note that the arrangement of messages among transmitters and receivers are similar to the generalized Z-interference networks in Fig. 12. These networks are a natural generalization of the so-called *Many-to-One Interference Channel* [31, 32] to the multi-message case. Hence, they are coined as the "Many-to-One Interference Networks". In these networks many *groups* of transmitters send information to their respective receivers (each group is concerned to one receiver) via a common media such that all receivers, except one, receive interference-free signals. In other words, only one of the receivers experiences interference, i.e., its received signal contains information regarding both desired and non-desired messages. For the Gaussian Many-to-One classical IC, a noisy interference regime was identified in [33] using the outer bounds derived based on the genie-aided techniques. Here, we identify such a regime for the many-to-one interference networks in Fig. 13, for both discrete and Gaussian networks. Specifically, consider a network with the following constraint:

$$\mathbb{P}_{Y_1 \dots Y_{K_2} | \mathbb{X}_1 \dots \mathbb{X}_{K_2}} = \mathbb{P}_{Y_1|\mathbb{X}_1} \mathbb{P}_{Y_2|\mathbb{X}_2} \mathbb{P}_{Y_3|\mathbb{X}_3} \dots \mathbb{P}_{Y_{K_2-1}|\mathbb{X}_{K_2-1}} \mathbb{P}_{Y_{K_2}|Y_1 \dots Y_{K_2-1} \mathbb{X}_{K_2}} \tag{76}$$

For the networks in (76), conditioned on $\mathbb{X}_{K_2}$, the output Y_{K_2} is a noisy (degraded) version of $Y_1, \dots, Y_{K_2-1}$. In the next theorem, we prove that for such networks, treating interference as noise is sum-rate optimal.

Theorem 4.3) *Consider the many-to-one interference network in Fig. 13. If the transition probability function satisfies the degradedness condition in (76), then the sum-rate capacity is given below:*

$$\mathcal{C}_{sum}^{MtO-IN_{deg}} = \max_{\overset{*}{\mathcal{P}}_{sum}^{MAIN}} \Big(I(\mathbb{X}_1; Y_1|Q) + I(\mathbb{X}_2; Y_2|Q) + \dots + I\big(\mathbb{X}_{K_2-1}; Y_{K_2-1}\big|Q\big) + I\big(\mathbb{X}_{K_2}; Y_{K_2}\big|Q\big) \Big) \tag{77}$$

where $\overset{*}{\mathcal{P}}_{sum}^{MAIN}$ is given as in Proposition 4.1.

Proof of Theorem 4.3) To achieve the bound (77), similar to the scenario of Theorem 4.2, each receiver decodes its own messages and treats all the other signals as noise. For the converse part, note that the degradedness condition (76) implies the Markov relations given in (66). Therefore, (67) constitutes an outer bound on the sum-rate capacity of the network. The proof is completed by the fact that for the many-to-one interference network (75) shown in Fig. 13, the expression (67) is reduced to (77). ■

Remark: It should be remarked that the degradedness condition (76) is in general different from (72), and the result of Theorem 4.3 cannot be deduced from Theorem 4.2. In fact, for all generalized Z-interference networks (71), where their transition probability function implies the Markov chains (66), treating interference as noise is sum-rate optimal.

Let us consider the special case of many-to-one classical IC which is derived from the network in Fig. 13 by setting $K_1 = K_2 = K$, i.e., there exists only one transmitter with respect to each receiver. The Gaussian channel is formulated as follows:

$$\begin{cases} Y_i = X_i + Z_i, \qquad i = 1, \dots, K-1 \\ Y_K = a_1 X_1 + a_2 X_2 + \cdots + a_K X_K + Z_K \end{cases} \tag{78}$$

where $Z_1, \dots, Z_K$ are independent Gaussian noise with zero means and unit variances. Also, the input signals are subject to power constraints: $\mathbb{E}[X_i^2] \leq P_i, i = 1, \dots, K$. In [33], the authors showed that if the channel gains satisfy:

$$\sum_{i=1}^{K-1} a_i^2 \leq 1 \tag{79}$$

then, treating interference as noise is sum-rate optimal. One can show that the channel with (79) is (stochastically) degraded in the sense of (76). Thus, its sum-rate capacity is also derived from Theorem 4.3.

4.1.3 Incorporation of Successive Decoding and Treating Interference as Noise

In Chapter 3, we proved that for all degraded interference networks (12), the successive decoding scheme is sum-rate optimal. Also, in previous subchapters, we introduced classes of interference networks for which treating interference as noise at the receivers achieves the sum-rate capacity. Now we intend to present interesting scenarios for which a combination of these two schemes is optimal. We identify classes of interference networks which are composed of two different sets of receivers so that to achieve their sum-rate capacity, the receivers of one set apply the successive decoding strategy while the receivers of the other set treat interference as noise and decode only their own messages.

Consider a class of MAINs shown in Fig. 11 with $K_2 = \eta_1 + \eta_2$, where η_1 and η_2 are two arbitrary natural numbers. Let the transition probability function of the network satisfies the following factorization:

$$\mathbb{P}_{Y_1\ldots Y_{\eta_1}Y_{\eta_1+1}\ldots Y_{\eta_1+\eta_2}|\mathbb{X}_1\ldots\mathbb{X}_{\eta_1}\mathbb{X}_{\eta_1+1}\ldots\mathbb{X}_{\eta_1+\eta_2}} = \mathbb{P}_{Y_1|\mathbb{X}_1}\mathbb{P}_{Y_2|\mathbb{X}_2}\ldots\mathbb{P}_{Y_{\eta_1-1}|\mathbb{X}_{\eta_1-1}}\mathbb{P}_{Y_{\eta_1}|Y_1\ldots Y_{\eta_1-1}\mathbb{X}_{\eta_1}} \times \mathbb{P}_{Y_{\eta_1+1}|Y_{\eta_1}\mathbb{X}_{\eta_1+1}\ldots\mathbb{X}_{\eta_1+\eta_2}}\mathbb{P}_{Y_{\eta_1+2}|Y_{\eta_1+1}}\ldots\mathbb{P}_{Y_{\eta_1+\eta_2-1}|Y_{\eta_1+\eta_2-2}}\mathbb{P}_{Y_{\eta_1+\eta_2}|Y_{\eta_1+\eta_2-1}} \tag{80}$$

Note that the networks in (80) is neither degraded in the sense of (12) nor satisfies the Z-networks' factorization (71). Therefore, they belong to none of the scenarios considered before. In fact, such networks can be considered as a combination of the degraded networks in (12) and the many-to-one networks in (76). In the next theorem, we establish the sum-rate capacity for these networks where we show that treating interference as noise at the receivers $Y_1, \ldots, Y_{\eta_1}$ and the successive decoding scheme at the receivers $Y_{\eta_1+1}, \ldots, Y_{\eta_1+\eta_2}$ is optimal.

Theorem 4.4) *For the MAIN shown in Fig. 11 with $K_2 = \eta_1 + \eta_2$ if the transition probability function of the network satisfies the degradedness condition in* (80)*, the sum-rate capacity is given by:*

$$\max_{\overset{*}{\mathcal{P}}_{sum}^{MAIN}} \begin{pmatrix} I(\mathbb{X}_1; Y_1|Q) + I(\mathbb{X}_2; Y_2|Q) + \cdots + I(\mathbb{X}_{\eta_1-1}; Y_{\eta_1-1}|Q) + I(\mathbb{X}_{\eta_1}; Y_{\eta_1}|Q) \\ +I(\mathbb{X}_{\eta_1+1}; Y_{\eta_1+1}|\mathbb{X}_{\eta_1+2}, \ldots, \mathbb{X}_{\eta_1+\eta_2-1}, \mathbb{X}_{\eta_1+\eta_2}, Q) + \cdots \\ +I(\mathbb{X}_{\eta_1+\eta_2-1}; Y_{\eta_1+\eta_2-1}|\mathbb{X}_{\eta_1+\eta_2}, Q) + I(\mathbb{X}_{\eta_1+\eta_2}; Y_{\eta_1+\eta_2}|Q) \end{pmatrix} \tag{81}$$

where $\overset{*}{\mathcal{P}}_{sum}^{MAIN}$ is given as in Proposition 4.1.

Proof of Theorem 4.4) The achievability is derived using a combination of treating interference as noise and successive decoding strategy: The messages $\mathbb{M} - \mathbb{M}^*$ are not sent, where $\mathbb{M}^*$ are those elements of $\mathbb{M}$ which are not a satellite for any other message (considering the plan of messages). The messages $\mathbb{M}^*$ are encoded separately using independent codewords. The receiver $Y_j, j = 1, \ldots, \eta_1$, jointly decodes its respective messages $\mathbb{M}_{Y_j}^*$ and treats all the other signals as noise. The decoding scheme at the receivers $Y_{\eta_1+1}, \ldots, Y_{\eta_1+\eta_2}$ is the successive decoding scheme presented for the degraded networks given in (12). The receiver $Y_{\eta_1+\eta_2}$ successively decodes its respective messages $\mathbb{M}_{Y_{\eta_1+\eta_2}}^*$. The partial sum-rate due to this step is

$$I\left(\mathbb{M}_{Y_{\eta_1+\eta_2}}^*; Y_{\eta_1+\eta_2}\middle|Q\right)$$

At the receiver $Y_{\eta_1+\eta_2-1}$, first the messages $\mathbb{M}^*_{Y_{\eta_1+\eta_2}}$ are successively decoded similar to the receiver $Y_{\eta_1+\eta_2}$; this step does not introduce any new rate cost because $Y_{\eta_1+\eta_2}$ is a degraded version of $Y_{\eta_1+\eta_2-1}$. Then, it successively decodes its respective messages $\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}}$ using the received sequence as well as the previously decoded codewords. For this step, successful decoding with the following partial sum-rate is achievable:

$$I\left(\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}};Y_{\eta_1+\eta_2-1}\middle|\mathbb{M}^*_{Y_{\eta_1+\eta_2}},Q\right)$$

This successive decoding strategy is repeated at other receivers step by step from the weaker receivers towards the stronger ones up to the receiver Y_{η_1+1}. This receiver also first decodes the messages $\mathbb{M}^*_{Y_{\eta_1+\eta_2}},\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}},\dots,\mathbb{M}^*_{Y_{\eta_1+2}}$ successively and then decodes its own messages $\mathbb{M}^*_{Y_{\eta_1+1}}$. The rate cost due to this step is

$$I\left(\mathbb{M}^*_{Y_{\eta_1+1}};Y_{\eta_1+1}\middle|\mathbb{M}^*_{Y_{\eta_1+2}},\dots,\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}},\mathbb{M}^*_{Y_{\eta_1+\eta_2}},Q\right)$$

Therefore, by this scheme the following sum-rate is achieved:

$$\max_{\{\text{PDFs in (65)}\}}\begin{pmatrix} I\left(\mathbb{M}^*_{Y_1};Y_1\middle|Q\right)+I\left(\mathbb{M}^*_{Y_2};Y_2\middle|Q\right)+\cdots+I\left(\mathbb{M}^*_{Y_{\eta_1-1}};Y_{\eta_1-1}\middle|Q\right)+I\left(\mathbb{M}^*_{Y_{\eta_1}};Y_{\eta_1}\middle|Q\right) \\ I\left(\mathbb{M}^*_{Y_{\eta_1+1}};Y_{\eta_1+1}\middle|\mathbb{M}^*_{Y_{\eta_1+2}},\dots,\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}},\mathbb{M}^*_{Y_{\eta_1+\eta_2}},Q\right)+\cdots \\ +I\left(\mathbb{M}^*_{Y_{\eta_1+\eta_2-1}};Y_{\eta_1+\eta_2-1}\middle|\mathbb{M}^*_{Y_{\eta_1+\eta_2}},Q\right)+I\left(\mathbb{M}^*_{Y_{\eta_1+\eta_2}};Y_{\eta_1+\eta_2}\middle|Q\right)\end{pmatrix} \quad (82)$$

Similar to the procedure in (70), one can show that (82) is equivalent to (81). For the converse part, note that the degradedness condition in (80) implies the Markov relations (66). Therefore, (67) constitutes an outer bound on the sum-rate capacity. Moreover, given the factorization (80), the expression in (67) is reduced to (81). The proof is thus complete. ■

In fact, many other scenarios could be identified for which the outer bound in (60) is sum-rate optimal. Essentially, this outer bound is optimal for the degraded networks in (12) for which the successive decoding strategy achieves the sum-rate capacity. The degradedness condition in (12) implies that a weaker receiver, given a stronger receiver, is statistically independent of all the input signals. Nonetheless, for certain networks with receivers statistically independent of some of the transmitters, such as those introduced in this subchapter, it is always possible to somewhat relax the crucial degradedness constraint (12) and derive situations where the outer bound (60) is still sum-rate optimal.

Chapter 5
Nested Cut-Set Bound

Let us reconsider the outer bound established in Theorem 4.1 on the sum-rate capacity of general interference networks. This bound is indeed derived by enhancing a general non-degraded network with artificial outputs to obtain a degraded network whose capacity region includes that of the origin network as a subset. The degraded structure (61) imposed on the network is in fact the fundamental key which enables us to establish the single-letter outer bound (60). It should be noted that although the outer bound (60) is given based on auxiliary random variables, one can easily re-derive another bound (possibly weaker) which is only expressed based on the input signals. The approach to this end is similar to what we derived in (44) (please notice how auxiliaries are removed in (44) step by step). The resultant outer bound would be similar to (64) given for the MAINs. Another fact is that although the outer bound (60) is established on the sum-rate capacity of the interference networks, one can derive similar bounds on partial sum-rate capacities as well. As a result, we can obtain a unified outer bound on the capacity region of all single-hop communication networks which is generally tighter that the standard cut-set bound. This bound is indeed given in an early draft of a part of this research due to 2012 [10]. In 2014, Shomorony and Avestimehr reported a general outer bound on the capacity region of deterministic interference networks [9]. One can verify that their outer bound is in fact a special case of the general outer bound established by the author in [10].

In this Chapter, we intend to go beyond interference networks and establish a general outer bound on the capacity region of all communication networks. Clearly, by taking insight from our results for degraded interference networks, we present an extension to the standard cut-set for general communication networks of arbitrary large sizes, which we refer to as nested cut-set bound. The derivation is in spirit identical to that of degraded interference networks. Given a general communication network topology, we apply a series of cuts to the network to partition its nodes into distinct subsets. Next, we collect the outputs in each subset and treat them totally as the output signal corresponding to that subset. Finally, we impose a degraded structure similar to (61) on the output signals of the subsets and then employ the proof technique developed for degraded interference networks to derive a single-letter outer bound. The details are given as follows.

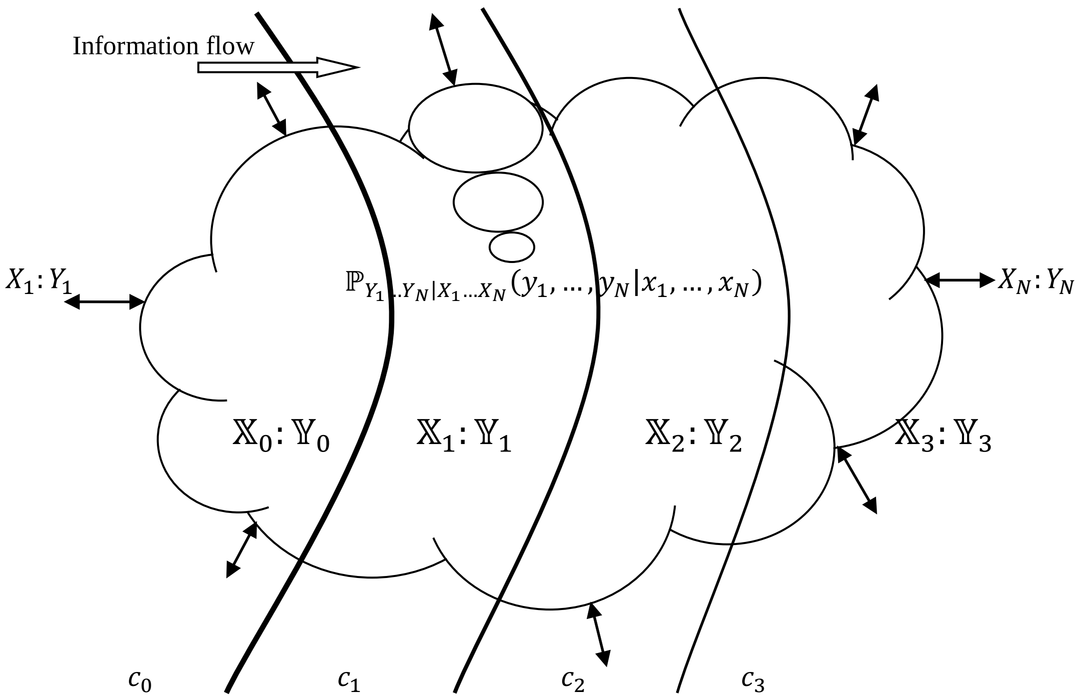

Figure 14. A series of three cuts applied to the general communication netwrok.

Consider the general communication network as described in Chapter 2.1. Given any integer r, a series of r cuts denoted by $\langle c_0, c_1, \ldots, c_r \rangle$ is a partitioning of the nodes of the general communication network into $r + 1$ distinct subsets so that i^{th} subset is denoted by $c_i, i \in [0, r]$. Figure 14 represents a series of three cuts.

For $i = 0,1, \ldots, r$, let define:

$$\mathbb{X}_i \triangleq \{X_k : X_k \in c_i\}, \qquad \mathbb{Y}_i \triangleq \{Y_k : Y_k \in c_i\}$$
$$\mathbb{M}_{\mathbb{X}_i} \triangleq \bigcup_{X_k \in \mathbb{X}_i} \mathbb{M}_{X_k}, \qquad \mathbb{M}_{\mathbb{Y}_i} \triangleq \bigcup_{Y_k \in \mathbb{Y}_i} \mathbb{M}_{Y_k} \tag{83}$$

Therefore, $\mathbb{X}_i$ and $\mathbb{Y}_i$ are respectively the collection of input and output signals of the nodes in c_i, and $\mathbb{M}_{\mathbb{X}_i}$ and $\mathbb{M}_{\mathbb{Y}_i}$ are respectively the set of messages which are transmitted and received by the nodes in c_i. Let also denote:

$$\mathbb{M}_i^{\langle c_0,c_1,\dots,c_r\rangle} \triangleq \mathbb{M}_{\mathbb{Y}_i} - \mathbb{M}_{\mathbb{X}_i} - \bigcup_{k>i} \mathbb{M}_{\mathbb{Y}_k} - \bigcup_{k>i} \mathbb{M}_{\mathbb{X}_k}, \qquad i = 0,1,\dots,r \tag{84}$$

The information rate which flows through the series of cuts $\langle c_0, c_1, \dots, c_r\rangle$ is defined as follows:

$$\boldsymbol{R}_{\langle c_0,c_1,\dots,c_r\rangle} \triangleq \sum_{k\in[1,r]} \sum_{\substack{i\ :\\ M_i\in\mathbb{M}_k^{\langle c_0,c_1,\dots,c_r\rangle}}} R_i \tag{85}$$

Now we are at the point to state our main result of this chapter as given in the following theorem.

Theorem 5.1) *Consider a general communication network with N nodes as depicted in Fig. 1. Assume that the network is cut by a series of r cuts denoted by $\langle c_0, c_1, \dots, c_r\rangle$. The information rate flowing through the cuts is bounded as follows:*

$$\boldsymbol{R}_{\langle c_0,c_1,\dots,c_r\rangle} \leq I(\mathbb{X}_0; \mathbb{Y}_1, \mathbb{Y}_2, \dots, \mathbb{Y}_r | \mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_r) + I(\mathbb{X}_1; \mathbb{Y}_2, \dots, \mathbb{Y}_r | \mathbb{X}_2, \dots, \mathbb{X}_r) + \cdots$$
$$+ I(\mathbb{X}_{r-2}; \mathbb{Y}_{r-1}, \mathbb{Y}_r | \mathbb{X}_{r-1}, \mathbb{X}_r) + I(\mathbb{X}_{r-1}; \mathbb{Y}_r | \mathbb{X}_r) \tag{86}$$

For some joint distributions $P_{X_1X_2\dots X_N}(x_1, x_2, \dots, x_N)$.

Proof of Theorem 5.1: As mentioned before, the proof is essentially similar to what we derived for degraded networks in Chapter 3. As it is clear from the bound (86), the key is the following degraded structure which we impose on the network using the series of cuts.

$$\mathbb{X}_0, \mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_r \to \mathbb{Y}_1, \mathbb{Y}_2, \dots, \mathbb{Y}_r \to \mathbb{Y}_2, \dots, \mathbb{Y}_r \to \cdots \to \mathbb{Y}_{r-1}, \mathbb{Y}_r \to \mathbb{Y}_r \tag{87}$$

Let describe the proof in details. Consider a length-n code $\mathfrak{C}^n(R_1, \dots, R_K)$ with vanishing average error probability for the network. Using Fano's inequality, we have:

$$\sum_{i\ :\ M_i\in\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}} R_i \leq \frac{1}{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_r^n, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r,n}$$
$$= \frac{1}{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_r^n \middle| \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r,n}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r,t}\middle|\mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r,n} \\
&\overset{(a)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}, \mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r,n} \\
&\le \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_r^{t-1}; \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \epsilon_{r,n} \\
&\le \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_r^{t-1}; \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \epsilon_{r,n} \\
&\overset{(b)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_r^{t-1}; \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \epsilon_{r,n} \\
&\overset{(c)}{\le} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_r^{t-1}; \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \epsilon_{r,n}
\end{aligned}
\tag{88}
$$

where $\epsilon_{r,n} \to 0$ as $n \to \infty$, equality (a) holds because $\mathbb{X}_{r,t}$ is a deterministic function of $\left(\mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{X}_r}\right)$, equality (b) holds because $\mathbb{M}_r^{\langle c_0,c_1,\dots,c_r\rangle} \cup \mathbb{M}_{\mathbb{X}_r}^n = \mathbb{M}_{\mathbb{Y}_r}^n \cup \mathbb{M}_{\mathbb{X}_r}^n$, and inequality (c) holds because adding information increases mutual information. Also, we have:

$$
\begin{aligned}
\sum_{i\,:\,M_i \in \mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}} R_i &\le \frac{1}{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r-1}^n, \mathbb{M}_{\mathbb{X}_{r-1}}\right) + \epsilon_{r-1,n} \\
&\le \frac{1}{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r-1}^n, \mathbb{Y}_r^n, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r-1,n} \\
&= \frac{1}{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r-1}^n, \mathbb{Y}_r^n\middle|\mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r-1,n} \\
&= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\middle|\mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r-1,n} \\
&\overset{(d)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r-1,n} \\
&\le \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{r-1}^{\langle c_0,c_1,\dots,c_r\rangle}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\middle|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_r^{t-1}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right) + \epsilon_{r-1,n}
\end{aligned}
$$

$$\overset{(e)}{=} \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\big|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\big) + \epsilon_{r-1,n}$$
$$\overset{(f)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{X}_{r-2,t}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\big|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\big) + \epsilon_{r-1,n} \tag{89}$$

where $\epsilon_{r-1,n} \to 0$ as $n \to \infty$, equality (d) holds because $\mathbb{X}_{r-1,t}$ and $\mathbb{X}_{r,t}$ are deterministic functions of $\big(\mathbb{Y}_{r-1}^{t-1}, \mathbb{M}_{\mathbb{X}_{r-1}}\big)$ and $\big(\mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{X}_{r}}\big)$, respectively, equality (e) holds because $\mathbb{M}_{r-1}^{\langle c_0,c_1,\ldots,c_r\rangle} \cup \mathbb{M}_{\mathbb{X}_{r-1}} \cup \mathbb{M}_{\mathbb{Y}_r} \cup \mathbb{M}_{\mathbb{X}_r} = \mathbb{M}_{\mathbb{Y}_{r-1}} \cup \mathbb{M}_{\mathbb{X}_{r-1}} \cup \mathbb{M}_{\mathbb{Y}_r} \cup \mathbb{M}_{\mathbb{X}_r}$, and inequality (f) holds because adding information increase mutual information. By the same approach, for $k = 1, \ldots, r$, one can derive:

$$\sum_{i\ :\ M_i \in \mathbb{M}_k^{\langle c_0,c_1,\ldots,c_r\rangle}} R_i \leq \Sigma_k + \epsilon_{k,n} \tag{90}$$

where $\epsilon_{k,n} \to 0$ as $n \to \infty$, and Σ_k is defined as follows:

$$\Sigma_k \triangleq$$
$$\frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{X}_{k-1,t}, \mathbb{M}_{\mathbb{Y}_k}, \mathbb{M}_{\mathbb{X}_k}, \mathbb{Y}_{k}^{t-1}; \mathbb{Y}_{k,t}, \ldots, \mathbb{Y}_{r,t}\big|\mathbb{X}_{k,t}, \ldots, \mathbb{X}_{r,t}, \mathbb{Y}_{k+1}^{t-1}, \ldots, \mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{Y}_{k+1}}, \mathbb{M}_{\mathbb{X}_{k+1}}, \ldots, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\big) \tag{91}$$

Then, we have:

$$\begin{aligned}
\Sigma_r + \Sigma_{r-1} &= \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{X}_{r-1,t}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r}^{t-1}; \mathbb{Y}_{r,t}\big|\mathbb{X}_{r,t}\big) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{X}_{r-2,t}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t}\big|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\big) \\
&= \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{X}_{r-1,t}; \mathbb{Y}_{r,t}\big|\mathbb{X}_{r,t}\big) + \frac{1}{n}\sum_{t=1}^{n} I\big(\mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r}^{t-1}; \mathbb{Y}_{r,t}\big|\mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\big)
\end{aligned}$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r}^{t-1}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t}; \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}} \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_{r}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$(92)$$

Also, we can derive:

$$\varSigma_r + \varSigma_{r-1} + \varSigma_{r-2}$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t}; \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}} \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r}^{t-1}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-3,t}, \mathbb{M}_{\mathbb{Y}_{r-2}}, \mathbb{M}_{\mathbb{X}_{r-2}}, \mathbb{Y}_{r-2}^{t-1}; \mathbb{Y}_{r-2,t}, \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \begin{array}{c} \mathbb{X}_{r-2,t}, \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_{r}^{t-1}, \\ \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r} \end{array}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t}; \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}} \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r}^{t-1}, \mathbb{Y}_{r-1}^{t-1}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-2,t}, \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-3,t}, \mathbb{M}_{\mathbb{Y}_{r-2}}, \mathbb{M}_{\mathbb{X}_{r-2}}, \mathbb{Y}_{r-2}^{t-1}; \mathbb{Y}_{r-2,t}, \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \begin{array}{c} \mathbb{X}_{r-2,t}, \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}, \mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_{r}^{t-1}, \\ \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r} \end{array}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t}; \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t}; \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} I\left(\begin{array}{c} \mathbb{X}_{r-3,t}, \mathbb{M}_{\mathbb{Y}_{r-2}}, \mathbb{M}_{\mathbb{X}_{r-2}}, \mathbb{M}_{\mathbb{Y}_{r-1}}, \mathbb{M}_{\mathbb{X}_{r-1}}, \\ \mathbb{M}_{\mathbb{Y}_r}, \mathbb{M}_{\mathbb{X}_r}, \mathbb{Y}_{r-2}^{t-1}, \mathbb{Y}_{r-1}^{t-1}, \mathbb{Y}_{r}^{t-1} \end{array}; \mathbb{Y}_{r-2,t}, \mathbb{Y}_{r-1,t}, \mathbb{Y}_{r,t} \middle| \mathbb{X}_{r-2,t}, \mathbb{X}_{r-1,t}, \mathbb{X}_{r,t}\right)$$

$$(93)$$

By continuing the steps (92) and (93), we can obtain:

$$
\begin{aligned}
&\Sigma_r + \Sigma_{r-1} + \Sigma_{r-2} + \cdots + \Sigma_1 \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t};\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t};\mathbb{Y}_{r-1,t},\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r-1,t},\mathbb{X}_{r,t}\right) + \cdots \\
&\qquad\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{1,t};\mathbb{Y}_{2,t},\dots,\mathbb{Y}_{r,t}\middle|\mathbb{X}_{2,t},\dots,\mathbb{X}_{r,t}\right) \\
&\qquad\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\begin{matrix}\mathbb{X}_{0,t},\mathbb{M}_{\mathbb{Y}_1},\mathbb{M}_{\mathbb{X}_1},\mathbb{M}_{\mathbb{Y}_2},\mathbb{M}_{\mathbb{X}_2},\dots,\\ \mathbb{M}_{\mathbb{Y}_r},\mathbb{M}_{\mathbb{X}_r},Y_1^{t-1},\dots,Y_{r-1}^{t-1},Y_r^{t-1}\end{matrix};\mathbb{Y}_{1,t},\dots,\mathbb{Y}_{r,t}\middle|\mathbb{X}_{1,t},\dots,\mathbb{X}_{r,t}\right) \\
&\quad \stackrel{(g)}{=} \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t};\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t};\mathbb{Y}_{r-1,t},\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r-1,t},\mathbb{X}_{r,t}\right) + \cdots \\
&\qquad\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{0,t};\mathbb{Y}_{1,t},\dots,\mathbb{Y}_{r,t}\middle|\mathbb{X}_{1,t},\dots,\mathbb{X}_{r,t}\right)
\end{aligned}
\tag{94}
$$

where equality (g) holds because due to the memoryless characteristic of the network we have:

$$
\frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{M}_{\mathbb{Y}_1},\mathbb{M}_{\mathbb{X}_1},\mathbb{M}_{\mathbb{Y}_2},\mathbb{M}_{\mathbb{X}_2},\dots,\mathbb{M}_{\mathbb{Y}_r},\mathbb{M}_{\mathbb{X}_r},Y_1^{t-1},\dots,Y_{r-1}^{t-1},Y_r^{t-1};\mathbb{Y}_{1,t},\dots,\mathbb{Y}_{r,t}\middle|\mathbb{X}_{0,t},\mathbb{X}_{1,t},\dots,\mathbb{X}_{r,t}\right) = 0
\tag{95}
$$

Thus, using (94), we conclude:

$$
\begin{aligned}
\boldsymbol{R}_{\langle c_0,c_1,\dots,c_r\rangle} &= \sum_{k\in[1,r]} \sum_{\substack{i\ :\\ M_i\in\mathbb{M}_k^{\langle c_0,c_1,\dots,c_r\rangle}}} R_i \leq \sum_{k=1}^{r} \Sigma_k + \epsilon_{k,n} \\
&\leq \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{0,t};\mathbb{Y}_{1,t},\dots,\mathbb{Y}_{r,t}\middle|\mathbb{X}_{1,t},\dots,\mathbb{X}_{r,t}\right) + \cdots + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-2,t};\mathbb{Y}_{r-1,t},\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r-1,t},\mathbb{X}_{r,t}\right) \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} I\left(\mathbb{X}_{r-1,t};\mathbb{Y}_{r,t}\middle|\mathbb{X}_{r,t}\right) + \sum_{k=1}^{r}\epsilon_{k,n}
\end{aligned}
\tag{96}
$$

The proof is now complete. ■

Remarks:

1. It is clear that the nested cut-set bound (86) coincides with the standard cut-set bound if we consider just one cut. For any network with more than one receiver, the nested cut-set bound would be strictly tighter than the cut-set bound.
2. By comparing the proofs of Theorems 3.1 and 5.1, we see that the derivations for the nested cut-set bound (86) are exactly similar to those of the sum-rate capacity for general degraded interference networks. However, for degraded networks, we keep the messages in final expression because they are needed to represent the sum-rate capacity. Instead for the nested cut-set bound, we derive the final expression based on the input signals which yields a simple computation of the bound.
3. Though the bound in (86) is given for all joint distributions on the input signals, for many networks where some inputs are statistically independent of each other, e.g., the interference channels, one can consider a more limited set of distributions. In this case, albeit one needs to a time-sharing parameter for characterizing the bound. Thus, the outer bound (64) is indeed the nested cut-set bound for MAINs. This shows that, unlike the standard cut-set bound, the nested cut-set bound is indeed optimal or sum-rate optimal for many large communication networks.
4. The nested cut-set bound holds also for the general Gaussian communication networks (6). Using the principle of "Gaussian maximizes the entropy", one can simply show that Gaussian inputs are always optimal for the bound. Therefore, it can be easily evaluated for any Gaussian network.

We conclude this chapter by providing an example on the nest cut-set bound. Consider a two-user interference channel with a relay as shown in Fig. 15 on the top of next page. In this network, two transmitters send independent messages to their corresponding receivers while being assisted by a relay. This scenario is called the interference-relay channel. For the interference-relay channel, the standard cut-set bound is given as follows:

$$\begin{aligned} R_1 &\leq \min\{I(X_1,X_r;Y_1|X_2,Q), I(X_1;Y_1,Y_r|X_r,X_2,Q)\} \\ R_2 &\leq \min\{I(X_2,X_r;Y_2|X_1,Q), I(X_2;Y_2,Y_r|X_r,X_1,Q)\} \\ R_1+R_2 &\leq \min\{I(X_1,X_2,X_r;Y_1,Y_2|Q), I(X_1,X_2;Y_1,Y_2,Y_r|X_r,Q)\} \end{aligned} \tag{97}$$

for some joint PDFs $P_Q P_{X_1 Q} P_{X_2|Q} P_{X_r|X_1 X_2 Q}$.

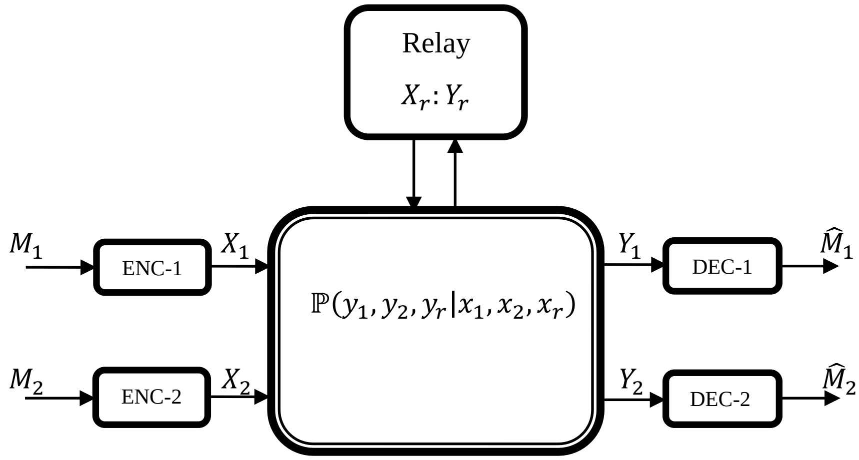

Figure 15. The interference-relay channel.

The nested cut-set bound includes the following constraints as well:

$$\begin{aligned}
R_1 + R_2 &\leq I(X_1, X_r; Y_1, Y_2|X_2, Q) + I(X_2; Y_2|Q) \\
R_1 + R_2 &\leq I(X_2, X_r; Y_1, Y_2|X_1, Q) + I(X_1; Y_1|Q) \\
R_1 + R_2 &\leq I(X_1; Y_1, Y_2, Y_r|X_2, X_r, Q) + I(X_2, X_r; Y_2|Q) \\
R_1 + R_2 &\leq I(X_2; Y_1, Y_2, Y_r|X_1, X_r, Q) + I(X_1, X_r; Y_1|Q) \\
R_1 + R_2 &\leq I(X_1; Y_1, Y_2, Y_r|X_2, X_r, Q) + I(X_2; Y_2, Y_r|X_r, Q) \\
R_1 + R_2 &\leq I(X_2; Y_1, Y_2, Y_r|X_1, X_r, Q) + I(X_1; Y_1, Y_r|X_r, Q)
\end{aligned} \tag{98}$$

The above constraints are indeed derived by applying two cuts to the network. It is clear that the nested cut-set bound is significantly tighter. The same conclusion can be derived for many other networks.

Chapter 6

Two-User Interference Channels with Conferencing Receivers

This chapter is divided into two parts. We present our results for the general two-user IC with cooperative users in Chapter 6.1. Then in Chapter 6.2, we specifically consider the Gaussian channel given in (9).

6.1 General Interference Channel with Conferencing Receivers

First of all, we present a novel outer bound for the general two-user IC with conferencing decoders.

Define $\mathfrak{R}_o^{IC\to CD}$ as the union of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ such that

$$\begin{aligned}
R_1 &\leq \min\{I(U,X_1;Y_1|Q) + D_{21}, I(X_1;Y_1|X_2,Q) + D_{21}\} \\
R_1 &\leq I(X_1;Y_1|Y_2,X_2,V,Q) + I(X_1;Y_2|X_2,Q) \\
R_1 &\leq I(X_1;Y_2|Y_1,X_2,V,Q) + I(X_1;Y_1|X_2,Q) \\
R_2 &\leq \min\{I(V,X_2;Y_2|Q) + D_{12}, I(X_2;Y_2|X_1,Q) + D_{12}\} \\
R_2 &\leq I(X_2;Y_2|Y_1,X_1,U,Q) + I(X_2;Y_1|X_1,Q) \\
R_2 &\leq I(X_2;Y_1|Y_2,X_1,U,Q) + I(X_2;Y_2|X_1,Q) \\
R_1 + R_2 &\leq I(X_1;Y_1|V,X_2,Q) + I(V,X_2;Y_2|Q) + D_{12} + D_{21} \\
R_1 + R_2 &\leq I(X_2;Y_2|U,X_1,Q) + I(U,X_1;Y_1|Q) + D_{12} + D_{21} \\
R_1 + R_2 &\leq I(X_1;Y_1|Y_2,X_2,V,Q) + I(X_1,X_2;Y_2|Q) + D_{12} \\
R_1 + R_2 &\leq I(X_2;Y_2|Y_1,X_1,U,Q) + I(X_1,X_2;Y_1|Q) + D_{21} \\
R_1 + R_2 &\leq I(X_1,X_2;Y_1,Y_2|Q)
\end{aligned} \tag{99}$$

for some joint PDFs $P_Q P_{X_1|Q} P_{X_2|Q} P_{UV|X_1 X_2 Q}$. The following theorem holds.

Theorem 6.1. *The set $\mathfrak{R}_o^{IC\to CD}$ constitutes an outer bound on the capacity region of the two-user IC with decoders connected by the conferencing links of capacities D_{12} and D_{21}, as shown in Fig. 3.*

Proof of Theorem 6.1: The proof is given in Appendix C. ■

Next, using the outer bound (99), we prove four capacity results for the IC with unidirectional conferencing between receivers. We highlight that a conferencing link (between receivers) may be utilized to provide one receiver with information about its corresponding signal or its non-corresponding signal (interference). Our following theorems reveal that both strategies can be helpful to achieve the capacity of the channel.

Theorem 6.2. *For the two-user IC with unidirectional conferencing between decoders, where* $D_{21} = 0$*, if*

$$\begin{gathered} I(X_1;Y_1|X_2) \leq I(X_1;Y_2|X_2) \quad \textit{for all } P_{X_1}P_{X_2} \\ X_2 \to Y_1, X_1 \to Y_2 \;\; \textit{(Markov chain)} \end{gathered} \tag{100}$$

then, the outer bound (99) is optimal. The capacity region is given by the union of all $(R_1, R_2) \in \mathbb{R}_+^2$ *such that:*

$$\begin{aligned} R_1 &\leq I(X_1;Y_1|X_2,Q), \\ R_2 &\leq min\{I(X_2;Y_2|X_1,Q) + D_{12}, I(X_2;Y_1|X_1,Q)\} \\ R_1 + R_2 &\leq min\{I(X_1,X_2;Y_2|Q) + D_{12}, I(X_1,X_2;Y_1|Q)\} \end{aligned} \tag{101}$$

for some joint PDFs $P_Q P_{X_1|Q} P_{X_2|Q}$.

Proof of Theorem 6.2: Let first prove the achievability of (101). Without loss of generality, assume that the time-sharing variable is null $Q \cong \emptyset$. We present a coding scheme in which both messages are decoded at both receivers. Consider the independent random variables M_1 and M_2 uniformly distributed over the sets $[1:2^{nR_1}]$ and $[1:2^{nR_2}]$, respectively. Partition the set $[1:2^{nR_2}]$ into $2^{nR_{12}}$ cells each containing $2^{n(R_2-R_{12})}$ elements, where $R_{12} = \min\{R_2, D_{12}\}$. Now label the cells by $c \in [1:2^{nR_{12}}]$ and the elements inside each cell by $\kappa \in \left[1:2^{n(R_2-R_{12})}\right]$. Accordingly, we have $c(M_2) = \theta$ if M_2 is inside the cell θ, and $\kappa(M_2) = \beta$ if M_2 is the β^{th} element of the cell that it belongs to.

Encoding at the transmitters is similar to a multiple access channel. For decoding, the first receiver decodes both messages M_1 and M_2, by exploring for codewords X_1^n and X_2^n which are jointly typical with its received sequence Y_1^n. This receiver then sends the cell index of the estimated message of the second transmitter, i.e. $c(\widehat{M}_2)$, to the receiver Y_2 by holding a $(D_{12}, 0)$-permissible conference. The second receiver applies a jointly typical decoder to decode the messages, with the caveat that the cell which M_2 belongs to is now known. Clearly, given $c(M_2)$, the second receiver detects the message M_1 and $\kappa(M_2)$ by exploring for codewords X_1^n and X_2^n which are jointly typical with its received sequence Y_2^n. One can easily show that under the conditions (100), this coding scheme yields the achievability of the rate region (101).

Next, using the outer bound (99), we show that under the conditions (100), the achievable rate region (101) is in fact optimal. Based on (99) for $D_{21} = 0$ we have:

$$
\begin{aligned}
R_2 &\leq I(X_2; Y_2 | Y_1, X_1, U, Q) + I(X_2; Y_1 | X_1, Q) \\
&\stackrel{a}{=} I(X_2; Y_1 | X_1, Q)
\end{aligned}
$$

$$
\begin{aligned}
R_1 + R_2 &\leq I(X_1; Y_1 | V, X_2, Q) + I(V, X_2; Y_2 | Q) + D_{12} \\
&\stackrel{b}{\leq} I(X_1; Y_2 | V, X_2, Q) + I(V, X_2; Y_2 | Q) + D_{12} \\
&= I(X_1, X_2; Y_2 | Q) + D_{12}
\end{aligned}
$$

$$
\begin{aligned}
R_1 + R_2 &\leq I(X_2; Y_2 | Y_1, X_1, U, Q) + I(X_1, X_2; Y_1 | Q) \\
&\stackrel{c}{=} I(X_1, X_2; Y_1 | Q)
\end{aligned}
$$

where equalities (a) and (c) are due to the second condition of (100) (given X_1, Y_2 is a degraded version of Y_1), and inequality (b) is due to the first condition of (100) (see [2, 34, 35]). Note that the other constraints of (101) are directly given by (99) when $D_{21} = 0$. The proof is thus complete. ■

Corollary 6.1. *Consider the following Gaussian IC with unidirectional conferencing between decoders (*$D_{21} = 0$*).*

$$
\begin{aligned}
Y_1 &= a_{11} X_1 + a_{12} X_2 + Z_1 \\
Y_2 &= a_{21} X_1 + a_{22} Y_1 + Z_2
\end{aligned}
\tag{102}
$$

where Z_1 *and* Z_2 *are independent unit-variance Gaussian noises. If* $(a_{11}^2 - a_{21}^2) \leq 2a_{11}a_{21}a_{22}$*, then the capacity region is given by (101).*

Proof of Corollary 6.1: First note that the channel (102) satisfies the second condition of (100) by definition. Moreover, one can easily see that for this channel the first condition of (100) is equivalent to $(a_{11}^2 - a_{21}^2) \leq 2a_{11}a_{21}a_{22}$. Therefore, we can apply the result of Theorem 6.2. ■

Based on Theorem 6.2, for the channel satisfying the conditions (100) the optimal scheme to achieve the capacity region is to decode both messages at both receivers and the optimal cooperation strategy is to provide one receiver with information about its corresponding signal via the conferencing link.

In fact, the conditions (100) could be interpreted as a strong interference regime for the IC with unidirectional cooperation between receivers. Note that if the channel satisfies (100), it will also satisfy the standard strong interference regime [27] as well.

Theorem 6.3. *For the two-user IC with unidirectional conferencing between decoders, where* $D_{21} = 0$, *if*

$$\begin{array}{cc} I(V;Y_2|X_2) \leq I(V;Y_1|X_2) & \textit{for all } P_{X_1}P_{X_2}P_{V|X_1X_2} \\ X_2 \to Y_1, X_1 \to Y_2 & \textit{(Markov chain)} \end{array} \tag{103}$$

then the outer bound (99) is sum-rate optimal and the sum-capacity is given by

$$\min_{P_Q P_{X_1|Q} P_{X_2|Q}} \left\{ \begin{array}{c} I(X_1;Y_1|X_2,Q) + I(X_2;Y_2|Q) + D_{12}, \\ I(X_1,X_2;Y_1|Q) \end{array} \right\} \tag{104}$$

Proof of Theorem 6.3: The coding scheme that achieves the sum-rate (104) is similar to that given in the proof of Theorem 6.2, except for the decoding of the second receiver. Here, the second receiver only decodes its own signal. Given $c(M_2)$, the second receiver detects $\kappa(M_2)$ by exploring for codewords X_2^n which are jointly typical with its received sequence Y_2^n. One can see that the sum-rate (104) is achieved by this scheme. Now consider the outer bound (99) where $D_{21} = 0$. Under the conditions (103), we have:

$$\begin{aligned} R_1 + R_2 &\leq I(X_1;Y_1|V,X_2,Q) + I(V,X_2;Y_2|Q) + D_{12} \\ &= I(X_1;Y_1|V,X_2,Q) + I(V;Y_2|X_2,Q) + I(X_2;Y_2|Q) + D_{12} \\ &\overset{a}{\leq} I(X_1;Y_1|V,X_2,Q) + I(V;Y_1|X_2,Q) + I(X_2;Y_2|Q) + D_{12} \\ &= I(X_1;Y_1|X_2,Q) + I(X_2;Y_2|Q) + D_{12} \end{aligned} \tag{105}$$

where inequality (a) is due to the first condition of (103). Moreover,

$$\begin{aligned} R_1 + R_2 &\leq I(X_2;Y_2|Y_1,X_1,U,Q) + I(X_1,X_2;Y_1|Q) \\ &\overset{b}{=} I(X_1,X_2;Y_1|Q) \end{aligned} \tag{106}$$

where equality (b) holds because of the second condition of (103), i.e., given X_1, Y_2 is a degraded version of Y_1, and thereby the first mutual information on the left side of (b) is zero. Therefore, (104) is in fact the sum-rate capacity of the channel and the proof is thus complete. ■

Corollary 6.2) *Consider the Gaussian IC given in (102) with unidirectional conferencing between decoders, where $D_{21} = 0$. If $(a_{11}^2 - a_{21}^2) \geq 2a_{11}a_{21}a_{22}$, then the sum-rate capacity is given by (104).*

Proof of Corollary 6.2: The Gaussian channel (102) satisfies the second condition of (103) by definition. Furthermore, for this channel the first condition of (103) is equivalent to $(a_{11}^2 - a_{21}^2) \geq 2a_{11}a_{21}a_{22}$. Thus, we can apply the result of Theorem 6.3. ■

According to Theorem 6.3, for the channel given in (103) the optimal scheme to achieve the sum-rate capacity is to decode interference at the receiver which is the source of the conferencing link, and to treat interference as noise at the receiver which is the destination of the conferencing link. Moreover, the optimal cooperation strategy is to provide the receiver that treats interference as noise with information about its corresponding signal via the conferencing link. The regime (103) could be indeed interpreted as a mixed interference regime for the IC with unidirectional cooperation between receivers. In fact, it is a special case of the mixed interference regime identified in [2, Th. 6] for the IC with no cooperation.

In the next theorem, we derive another mixed interference regime for the channel where, unlike Theorem 6.3, the optimal scheme to achieve the sum-capacity is to treat interference as noise at the receiver which is the source of the conferencing link and to decode interference at the one which is the destination of the conferencing link; also, the optimal cooperation strategy is to provide the receiver that decodes interference with information about its non-corresponding signal (interference) via the conferencing link.

Theorem 6.4) *For the two-user IC with unidirectional conferencing between receivers, where $D_{21} = 0$, if*

$$\begin{gathered} I(X_1;Y_1|X_2) \leq I(X_1;Y_2|X_2) \quad \text{for all } P_{X_1}P_{X_2} \\ X_2 \to Y_2, X_1 \to Y_1 \quad \text{(Markov chain)} \end{gathered} \tag{107}$$

then, the outer bound (99) is sum-rate optimal and the sum-rate capacity is given by

$$\min_{P_Q P_{X_1|Q} P_{X_2|Q}} \left\{ \begin{matrix} I(X_2;Y_2|X_1,Q) + I(X_1;Y_1|Q), \\ I(X_1,X_2;Y_2|Q) + D_{12} \end{matrix} \right\} \tag{108}$$

Proof of Theorem 6.4: The achievability of (108) is indeed derived by treating interference as noise at the first receiver and decoding interference at the second receiver. Moreover, the conferencing link is used to provide information about the interference for the second receiver. Let assume $Q \equiv \emptyset$. Consider two independent messages M_1 and M_2, uniformly distributed over the sets $[1:2^{nR_1}]$ and $[1:2^{nR_2}]$, respectively. Partition the set $[1:2^{nR_1}]$ into $2^{nR_{12}}$ cells each containing $2^{n(R_1-R_{12})}$ elements, where $R_{12} = \min\{R_1, D_{12}\}$. Now label the cells by $c \in [1:2^{nR_{12}}]$ and the elements inside each cell by $\kappa \in \left[1:2^{n(R_1-R_{12})}\right]$. Accordingly, we have $c(M_1) = \alpha$ if M_1 is inside the cell α, and $\kappa(M_1) = \beta$ if M_1 is β^{th} element of the cell that it belongs to.

Encoding at the transmitters is similar to a multiple access channel. For decoding, the first receiver simply decodes its own message by exploring for codewords X_1^n which are jointly typical with its received sequence Y_1^n. This receiver then sends the cell index of the estimated message, i.e. $c(\widehat{M}_1)$, to the second receiver by holding a $(D_{12},0)$-permissible conference. The second receiver applies a jointly typical decoder to decode both messages with the caveat that the cell index which M_1 belongs to is known. Clearly, given $c(M_1)$, the second receiver detects $\kappa(M_1)$ and M_2 by exploring for codewords X_1^n and X_2^n which are jointly typical with its received sequence Y_2^n. One can easily show that this scheme yields the achievable sum-rate (108).

Next using our outer bound (99) we prove that under the conditions (107), the sum-rate capacity of the channel is bounded by (108). Based on (99), when $D_{21} = 0$, we have:

$$\begin{aligned}
R_1 + R_2 &\leq I(X_2;Y_2|Y_1,X_1,U,Q) + I(X_1,X_2;Y_1|Q) \\
&= I(X_2;Y_1,Y_2|X_1,U,Q) - I(X_2;Y_1|X_1,U,Q) + I(X_2;Y_1|X_1,Q) + I(X_1;Y_1|Q) \\
&= I(X_2;Y_1,Y_2|X_1,U,Q) + I(U;Y_1|X_1,Q) + I(X_1;Y_1|Q) \\
&= I(X_2;Y_2|X_1,U,Q) + I(X_2;Y_1|Y_2,X_1,U,Q) + I(U;Y_1|X_1,Q) + I(X_1;Y_1|Q) \\
&\overset{a}{=} I(X_2;Y_2|X_1,U,Q) + I(U;Y_1|X_1,Q) + I(X_1;Y_1|Q) \\
&\overset{b}{\leq} I(X_2;Y_2|X_1,U,Q) + I(U;Y_2|X_1,Q) + I(X_1;Y_1|Q) \\
&= I(X_2;Y_2|X_1,Q) + I(X_1;Y_1|Q)
\end{aligned}$$

where equality (a) holds because by the Markov chain given in (107), the second mutual information on the left side of (a) is zero; similarly, inequality (b) holds because the Markov chain in (107) implies that $I(U;Y_1|X_1,Q) \leq I(U;Y_2|X_1,Q)$. Moreover, we have:

$$
\begin{aligned}
R_1 + R_2 &\leq I(X_1;Y_1|V,X_2,Q) + I(V,X_2;Y_2|Q) + D_{12} \\
&\overset{c}{\leq} I(X_1;Y_2|V,X_2,Q) + I(V,X_2;Y_2|Q) + D_{12} = I(X_1,X_2;Y_2|Q) + D_{12}
\end{aligned}
$$

where inequality (c) is due to the first condition of (107), (see [2, 34]). The proof is thus complete. ■

Corollary 6.3. *Consider the following Gaussian IC with unidirectional conferencing between decoders ($D_{21} = 0$).*

$$
\begin{aligned}
Y_1 &= a_{11}X_1 + a_{12}Y_2 + Z_1 \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + Z_2
\end{aligned}
\tag{109}
$$

where Z_1 and Z_2 are independent unit-variance Gaussian noises. If $(a_{21}^2 - a_{11}^2) \geq 2a_{11}a_{21}a_{12}$, then the sum-rate capacity is given by (108).

Proof of Corollary 6.3: The Gaussian channel (109) satisfies the second condition of (107) by definition. Moreover, for this channel the first condition of (107) is equivalent to $(a_{21}^2 - a_{11}^2) \geq 2a_{11}a_{21}a_{12}$. Thereby, we can apply the result of Theorem 6.4. ■

It should be remarked that although for the IC with conferencing receivers the mixed interference regimes (103) and (107) are different, for the channel with no cooperation both of them fall into the mixed interference regime identified in [2, Th.6][3].

Finally, we characterize the capacity region of the one-sided IC with a unidirectional conferencing link from the non-interfered receiver to the interfered one in the strong interference regime. The result is given in the last theorem of this subchapter.

Theorem 6.5. *Consider the two-user one-sided IC where $\mathbb{P}(y_1,y_2|x_1,x_2) = \mathbb{P}(y_1|x_1)\mathbb{P}(y_2|x_1,x_2)$. For the channel with unidirectional conferencing between receivers, where Y_1 is connected to Y_2 by a conferencing link of capacity D_{12}, if*

$$
I(X_1;Y_1|X_2) \leq I(X_1;Y_2|X_2) \quad \text{for all } P_{X_1}P_{X_2}
\tag{110}
$$

[3] Note that to see the regime (107) is a special case of the mixed interference regime identified in [2, Th.6], one need to exchange the indices 1 and 2 in (107) first.

then the outer bound (99) is optimal and the capacity region is given by the union of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ *such that*

$$\begin{aligned} R_1 &\leq I(X_1; Y_1|Q), \\ R_2 &\leq I(X_2; Y_2|X_1, Q) \\ R_1 + R_2 &\leq I(X_1, X_2; Y_2|Q) + D_{12} \end{aligned} \tag{111}$$

for some joint PDFs $P_Q P_{X_1|Q} P_{X_2|Q}$.

Proof of Theorem 6.5: The achievability proof is similar to the one presented in Theorem 6.4. The first receiver simply decodes its own message (as it does not perceive interference at all) while the second receiver jointly decodes both messages. The conferencing link is utilized to provide information about the interference (non-desired signal) for the second receiver. For the one-sided channel, this scheme achieves the rate region (111). Then we prove the converse part. Based on (99), when $D_{21} = 0$, we have:

$$R_1 \leq I(X_1; Y_1|X_2, Q) = I(X_1; Y_1|Q)$$

$$\begin{aligned} R_2 &\leq I(X_2; Y_1|Y_2, X_1, U, Q) + I(X_2; Y_2|X_1, Q) \\ &\overset{a}{=} I(X_2; Y_2|X_1, Q) \end{aligned}$$

$$\begin{aligned} R_1 + R_2 &\leq I(X_1; Y_1|V, X_2, Q) + I(V, X_2; Y_2|Q) + D_{12} \\ &\overset{b}{\leq} I(X_1; Y_2|V, X_2, Q) + I(V, X_2; Y_2|Q) + D_{12} \\ &= I(X_1, X_2; Y_2|Q) + D_{12} \end{aligned}$$

where (a) holds because for the one-sided IC, the first mutual information on the left side of (a) is zero; the inequality (b) is due to the condition (110). Thus, the proof is complete. ■

Corollary 6.4. *Consider the Gaussian one-sided IC which is given by* $a_{12} = 0$ *in (9). If* $|a_{21}| \geq |a_{11}|$, *then the capacity region of the channel with a unidirectional conferencing link from the non-interfered receiver to the interfered one is given by (111). This recovers a result of [25, Th.1].*

Let us summarize. For the scenarios considered in Theorems 6.2 and 6.3, the conferencing link is optimally utilized to provide a receiver with information about its corresponding signal, while for those considered in Theorems 6.4 and 6.5, it is optimally utilized to provide a receiver with information about its non-corresponding signal (interference signal). This demonstrates that, depending on the statistics of the channel, any of these strategies can be helpful to achieve the capacity region. In general, one may consider a combination of them as well. Clearly, one may design achievability schemes where the conferencing link is utilized to provide a receiver with information about its corresponding signal and its non-corresponding signal (interference) simultaneously. Indeed, it would be an interesting problem for future study to analyze the performance of such complex schemes.

6.2 General Interference Channel with Conferencing Receivers

The general outer bound (99) indeed holds for the Gaussian IC (9) with conferencing receivers as well (the input power constraints should also be considered when evaluating the bound). In this subchapter, we show that for the Gaussian channel, one can make the outer bound (99) tighter by establishing additional constraints on the information rates using genie-aided techniques. As a result, we obtain an outer bound which is strictly tighter than previous ones for all channel parameters.

Consider the two-user Gaussian IC in (9) with decoders connected by conferencing links of capacities D_{12} and D_{21}. Define genie signals $G_1, G_2, \tilde{G}_1$, and $\tilde{G}_2$ as follows:

$$\begin{aligned} G_1 &\triangleq a_{21}X_1 + Z_2 \\ G_2 &\triangleq a_{12}X_2 + Z_1 \\ \tilde{G}_1 &\triangleq a_{21}X_1 + \tilde{Z}_2 \\ \tilde{G}_2 &\triangleq a_{12}X_2 + \tilde{Z}_1 \end{aligned} \tag{112}$$

where $\tilde{Z}_1$ and $\tilde{Z}_2$ are unit-variance Gaussian noises independent of other variables. Let $\mathfrak{R}_{o,\langle UV \rangle}^{GIC\to CD}$ denote the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ which satisfy the constraints (99) as well as the following:

$$\begin{aligned} R_1 + R_2 &\le I(X_1, X_2; Y_1 | G_1, Q) + I(X_1, X_2; Y_2 | G_2, Q) + D_{12} + D_{21} \\ 2R_1 + R_2 &\le I(X_1; Y_1 | V, X_2, Q) + I(V, X_2; Y_2 | G_2, Q) + I(X_1, X_2; Y_1 | Q) + D_{12} + 2D_{21} \\ R_1 + 2R_2 &\le I(X_2; Y_2 | U, X_1, Q) + I(U, X_1; Y_1 | G_1, Q) + I(X_1, X_2; Y_2 | Q) + 2D_{12} + D_{21} \\ 2R_1 + R_2 &\le I(X_1; Y_1, Y_2 | V, X_2, Q) + I(V, X_2; Y_2 | G_2, Q) + I(X_1, X_2; Y_1 | Q) + D_{12} + D_{21} \\ R_1 + 2R_2 &\le I(X_2; Y_1, Y_2 | U, X_1, Q) + I(U, X_1; Y_1 | G_1, Q) + I(X_1, X_2; Y_2 | Q) + D_{12} + D_{21} \\ 2R_1 + R_2 &\le I\left(X_1, X_2; Y_1, Y_2 \middle| \tilde{G}_2, Q\right) + I(X_1, X_2; Y_1 | Q) + D_{21} \\ R_1 + 2R_2 &\le I\left(X_1, X_2; Y_1, Y_2 \middle| \tilde{G}_1, Q\right) + I(X_1, X_2; Y_2 | Q) + D_{12} \end{aligned} \tag{113}$$

for some joint PDFs $P_Q P_{X_1|Q} P_{X_2|Q} P_{UV|X_1X_2Q}$. The following theorem holds.

Theorem 6.6) *The set $\mathfrak{R}_{o,\langle UV\rangle}^{GIC\to CD}$ constitutes an outer bound on the capacity region of the two-user Gaussian IC (9) with conferencing decoders.*

Proof of Theorem 6.6) The proof is given in Appendix D. ■

We next present an explicit characterization of the outer bound given in Theorem 6.6. For this purpose, we indeed apply several novel techniques to optimize the bound over its auxiliary random variables. The result is given in the following theorem.

Theorem 6.7) *Let $\mathfrak{R}_{o}^{GIC\to CD}$ denote the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ which satisfy the constraints* (114) *given below for some $\alpha, \beta \in [0,1]$. The set $\mathfrak{R}_{o}^{GIC\to CD}$ constitutes an outer bound on the capacity region of the Gaussian IC* (9) *with decoders connected by the conferencing links of capacities D_{12} and D_{21}, as shown in Fig. 3.*

$$R_1 \leq \min\left\{\psi\left(\frac{\mathrm{SNR}_1+(1-\alpha)\mathrm{INR}_1}{\alpha\,\mathrm{INR}_1+1}\right)+D_{21}, \psi(\mathrm{SNR}_1)+D_{21}\right\} \tag{1-114}$$

$$R_1 \leq \min\left\{\psi\left(\frac{\beta\,\mathrm{SNR}_1}{\beta\,\mathrm{INR}_2+1}\right)+\psi(\mathrm{INR}_2), \psi\left(\frac{\beta\,\mathrm{INR}_2}{\beta\,\mathrm{SNR}_1+1}\right)+\psi(\mathrm{SNR}_1)\right\} \tag{2-114}$$

$$R_2 \leq \min\left\{\psi\left(\frac{(1-\beta)\mathrm{INR}_2+\mathrm{SNR}_2}{\beta\,\mathrm{INR}_2+1}\right)+D_{12}, \psi(\mathrm{SNR}_2)+D_{12}\right\} \tag{3-114}$$

$$R_2 \leq \min\left\{\psi\left(\frac{\alpha\,\mathrm{SNR}_2}{\alpha\,\mathrm{INR}_1+1}\right)+\psi(\mathrm{INR}_1), \psi\left(\frac{\alpha\,\mathrm{INR}_1}{\alpha\,\mathrm{SNR}_2+1}\right)+\psi(\mathrm{SNR}_2)\right\} \tag{4-114}$$

$$\begin{aligned}R_1+R_2 \leq &\left(\psi(\beta\,\mathrm{SNR}_1)+\psi\left(\frac{(1-\beta)\mathrm{INR}_2+\mathrm{SNR}_2}{\beta\,\mathrm{INR}_2+1}\right)\right)\mathbb{1}(|a_{21}|<|a_{11}|)\\ &+\psi(\mathrm{INR}_2+\mathrm{SNR}_2)\mathbb{1}(|a_{21}|\geq|a_{11}|)+D_{12}+D_{21}\end{aligned} \tag{5-114}$$

$$\begin{aligned}R_1+R_2 \leq &\left(\psi(\alpha\,\mathrm{SNR}_2)+\psi\left(\frac{\mathrm{SNR}_1+(1-\alpha)\mathrm{INR}_1}{\alpha\,\mathrm{INR}_1+1}\right)\right)\mathbb{1}(|a_{12}|<|a_{22}|)\\ &+\psi(\mathrm{SNR}_1+\mathrm{INR}_1)\mathbb{1}(|a_{12}|\geq|a_{22}|)+D_{12}+D_{21}\end{aligned} \tag{6-114}$$

$$R_1+R_2 \leq \psi\left(\frac{\beta\,\mathrm{SNR}_1}{\beta\,\mathrm{INR}_2+1}\right)+\psi(\mathrm{INR}_2+\mathrm{SNR}_2)+D_{12} \tag{7-114}$$

$$R_1 + R_2 \leq \psi\left(\frac{\alpha\,\mathrm{SNR}_2}{\alpha\,\mathrm{INR}_1+1}\right) + \psi(\mathrm{SNR}_1 + \mathrm{INR}_1) + D_{21}$$

(8-114)

$$R_1 + R_2 \leq \psi(\mathrm{SNR}_1 + \mathrm{INR}_1 + \mathrm{INR}_2 + \mathrm{SNR}_2 + |a_{11}a_{22} - a_{12}a_{21}|^2 P_1 P_2)$$

(9-114)

$$R_1 + R_2 \leq \psi\left(\mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{\mathrm{INR}_2+1}\right) + \psi\left(\mathrm{INR}_2 + \frac{\mathrm{SNR}_2}{\mathrm{INR}_1+1}\right) + D_{12} + D_{21}$$

(10-114)

$$2R_1 + R_2 \leq \big(\psi(\beta\,\mathrm{SNR}_1) - \psi(\beta\,\mathrm{INR}_2)\big)\mathbb{1}(|a_{21}| < |a_{11}|) + \psi\left(\mathrm{INR}_2 + \frac{\mathrm{SNR}_2}{\mathrm{INR}_1+1}\right)$$
$$+\psi(\mathrm{SNR}_1 + \mathrm{INR}_1) + D_{12} + 2D_{21}$$

(11-114)

$$R_1 + 2R_2 \leq \big(\psi(\alpha\,\mathrm{SNR}_2) - \psi(\alpha\,\mathrm{INR}_1)\big)\mathbb{1}(|a_{12}| < |a_{22}|) + \psi\left(\mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{\mathrm{INR}_2+1}\right)$$
$$+\psi(\mathrm{INR}_2 + \mathrm{SNR}_2) + 2D_{12} + D_{21}$$

(12-114)

$$2R_1 + R_2 \leq \psi\big(\beta\,(\mathrm{SNR}_1 + \mathrm{INR}_2)\big) + \psi\left(\frac{(1-\beta)\mathrm{INR}_2}{1+\beta\,\mathrm{INR}_2} + \frac{\mathrm{SNR}_2}{(\mathrm{INR}_1+1)(1+\beta\,\mathrm{INR}_2)}\right)$$
$$+\psi(\mathrm{SNR}_1 + \mathrm{INR}_1) + D_{12} + D_{21}$$

(13-114)

$$R_1 + 2R_2 \leq \psi\big(\alpha(\mathrm{INR}_1 + \mathrm{SNR}_2)\big) + \psi\left(\frac{(1-\alpha)\mathrm{INR}_1}{1+\alpha\,\mathrm{INR}_1} + \frac{\mathrm{SNR}_1}{(\mathrm{INR}_2+1)(1+\alpha\,\mathrm{INR}_1)}\right)$$
$$+\psi(\mathrm{INR}_2 + \mathrm{SNR}_2) + D_{12} + D_{21}$$

(14-114)

$$2R_1 + R_2 \leq \psi\left(\mathrm{SNR}_1 + \frac{\mathrm{INR}_1}{1+\mathrm{INR}_1} + \mathrm{INR}_2 + \frac{\mathrm{SNR}_2}{1+\mathrm{INR}_2} + \frac{|a_{11}a_{22}-a_{12}a_{21}|^2 P_1 P_2}{1+\mathrm{INR}_1}\right)$$
$$+\psi(\mathrm{SNR}_1 + \mathrm{INR}_1) + D_{21}$$

(15-114)

$$R_1 + 2R_2 \leq \psi\left(\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_2} + \mathrm{INR}_1 + \frac{\mathrm{INR}_2}{1+\mathrm{INR}_2} + \mathrm{SNR}_2 + \frac{|a_{11}a_{22}-a_{12}a_{21}|^2 P_1 P_2}{1+\mathrm{INR}_2}\right)$$
$$+\psi(\mathrm{INR}_2 + \mathrm{SNR}_2) + D_{12}$$

(16-114)

(114)

Proof of Theorem 6.7) We need to optimize the outer bound established in Theorem 6.6 over the auxiliary random variables U and V, which is indeed a complicated problem. To solve it, we apply novel techniques including several subtle applications of the entropy power inequality. Let present our approach. First note that some of the mutual information functions given in (99) and (113) can be re-written as follows:

$$\begin{aligned}
I(X_2;Y_2|Y_1,X_1,U,Q) &= I(X_2;Y_1,Y_2|U,X_1,Q) - I(X_2;Y_1|U,X_1,Q)\\
I(X_2;Y_1|Y_2,X_1,U,Q) &= I(X_2;Y_1,Y_2|U,X_1,Q) - I(X_2;Y_2|U,X_1,Q)\\
I(X_1;Y_1|Y_2,X_2,V,Q) &= I(X_1;Y_1,Y_2|V,X_2,Q) - I(X_1;Y_2|V,X_2,Q)\\
I(X_1;Y_2|Y_1,X_2,V,Q) &= I(X_1;Y_1,Y_2|V,X_2,Q) - I(X_1;Y_1|V,X_2,Q)
\end{aligned} \tag{115}$$

In general, it is difficult to directly treat expressions such as $I(X_1;Y_1,Y_2|V,X_2,Q)$ or $I(X_2;Y_1,Y_2|U,X_1,Q)$. To make the problem tractable, we apply the following technique. Let define two new outputs $\hat{Y}_1$ and $\hat{Y}_2$ as follows:

$$\begin{aligned}
\hat{Y}_1 &\triangleq \frac{a_{12}Y_1 + a_{22}Y_2}{|a_{12}|^2 + |a_{22}|^2} = \frac{a_{11}a_{12} + a_{21}a_{22}}{|a_{12}|^2 + |a_{22}|^2}X_1 + X_2 + \hat{Z}_1\\
\hat{Y}_2 &\triangleq \frac{a_{11}Y_1 + a_{21}Y_2}{|a_{11}|^2 + |a_{21}|^2} = X_1 + \frac{a_{11}a_{12} + a_{21}a_{22}}{|a_{11}|^2 + |a_{21}|^2}X_2 + \hat{Z}_2
\end{aligned} \tag{116}$$

where $\hat{Z}_1$ and $\hat{Z}_2$ are given as:

$$\begin{aligned}
\hat{Z}_1 &\triangleq \frac{a_{12}Z_1 + a_{22}Z_2}{|a_{12}|^2 + |a_{22}|^2}\\
\hat{Z}_2 &\triangleq \frac{a_{11}Z_1 + a_{21}Z_2}{|a_{11}|^2 + |a_{21}|^2}
\end{aligned} \tag{117}$$

It is clear that the mapping from (Y_1,Y_2) to $(\hat{Y}_1,Y_2)$ and also to $(Y_1,\hat{Y}_2)$ is one-to-one. Now we have:

$$\begin{aligned}
Y_1 &= a_{11}\hat{Y}_2 + \frac{a_{21}(a_{12}a_{21} - a_{11}a_{22})}{|a_{11}|^2 + |a_{21}|^2}X_2 + \bar{\bar{Z}}_1\\
Y_2 &= a_{22}\hat{Y}_1 + \frac{a_{12}(a_{12}a_{21} - a_{11}a_{22})}{|a_{12}|^2 + |a_{22}|^2}X_1 + \bar{\bar{Z}}_2
\end{aligned} \tag{118}$$

where

$$\bar{\bar{Z}}_1 \triangleq \frac{a_{21}(a_{21}Z_1 - a_{11}Z_2)}{|a_{11}|^2 + |a_{21}|^2}$$

$$\bar{\bar{Z}}_2 \triangleq \frac{a_{12}(a_{12}Z_2 - a_{22}Z_1)}{|a_{12}|^2 + |a_{22}|^2} \tag{119}$$

One can easily check that $\bar{\bar{Z}}_1$ is independent of $\hat{Z}_1$ and also $\bar{\bar{Z}}_2$ is independent of $\hat{Z}_2$. Therefore, the following equalities hold:

$$I(X_2; Y_1, Y_2|U, X_1, Q) = I\big(X_2; \hat{Y}_1, Y_2\big|U, X_1, Q\big) = I\big(X_2; \hat{Y}_1\big|U, X_1, Q\big)$$
$$I(X_1; Y_1, Y_2|V, X_2, Q) = I\big(X_1; Y_1, \hat{Y}_2\big|V, X_2, Q\big) = I\big(X_1; \hat{Y}_2\big|V, X_2, Q\big) \tag{120}$$

for any arbitrary input distributions. Next fix a distribution $P_Q P_{X_1|Q} P_{X_2|Q} P_{UV|X_1X_2Q}$ with $\mathbb{E}\big[X_i^2\big] \leq P_i, i = 1,2$. In what follows, we present the optimization for the auxiliary random variable U. The optimization over V is derived symmetrically, and therefore we do not present the details to avoid repetition.

Let divide the problem into two different cases. First consider the channel with weak interference at the first receiver where

$$|a_{12}| < |a_{22}| \tag{121}$$

It is clear that:

$$\frac{1}{2}\log 2\pi e = H(Y_2|U, X_1, X_2, Q) \leq H(Y_2|U, X_1, Q) = H(a_{22}X_2 + Z_2|U, X_1, Q)$$
$$\leq H(a_{22}X_2 + Z_2|Q) \leq \frac{1}{2}\log 2\pi e(|a_{22}|^2 P_2 + 1) \tag{122}$$

Comparing the two sides of (122), one can deduce that there is $\alpha \in [0,1]$ such that:

$$H(Y_2|U, X_1, Q) = H(a_{22}X_2 + Z_2|U, X_1, Q)$$
$$= \frac{1}{2}\log 2\pi e(|a_{22}|^2 \alpha P_2 + 1) \tag{123}$$

Then by considering (123) and also (115) and (120), one can easily verify that the optimization is equivalent to maximize $H(\hat{Y}_1|U, X_1, Q)$ and minimize $H(Y_1|U, X_1, Q)$, simultaneously. For this purpose, we apply the entropy power inequality. This inequality implies that for any arbitrary random variables X, Z, and W, where X and Z are independent conditioned on W, the following holds:

$$\exp\big(2H(X+Z|W)\big) \geq \exp\big(2H(X|W)\big) + \exp\big(2H(Z|W)\big) \tag{124}$$

Therefore, assuming that $H(Z|W)$ is fixed, given $H(X+Z|W)$, one can derive an upper bound on $H(X|W)$, and given $H(X|W)$, one can derive a lower bound on $H(X+Z|W)$. This fact is the essence of our arguments in what follows.

Let $\hat{Z}_1^*$ be a Gaussian random variable, independent of all other variables, with zero mean and a variance equal to $\frac{|a_{12}|^2}{|a_{12}|^2+|a_{22}|^2}$. One can write:

$$\begin{aligned}
&H(\hat{Y}_1|U, X_1, Q)\\
&\quad = H\left(\frac{a_{12}Y_1 + a_{22}Y_2}{|a_{12}|^2+|a_{22}|^2}\middle|U, X_1, Q\right)\\
&\quad = H(X_2 + \hat{Z}_1|U, X_1, Q)\\
&\quad = H(a_{22}X_2 + a_{22}\hat{Z}_1|U, X_1, Q) - \frac{1}{2}\log|a_{22}|^2\\
&\quad \overset{a}{\leq} \frac{1}{2}\log\left(\exp\left(2H(a_{22}X_2 + a_{22}\hat{Z}_1 + \hat{Z}_1^*|U, X_1, Q)\right) - \exp\left(2H(\hat{Z}_1^*|U, X_1, Q)\right)\right) - \frac{1}{2}\log|a_{22}|^2\\
&\quad \overset{b}{=} \frac{1}{2}\log\left(\exp(2H(a_{22}X_2 + Z_2|U, X_1, Q)) - \exp\left(2H(\hat{Z}_1^*)\right)\right) - \frac{1}{2}\log|a_{22}|^2\\
&\quad \overset{c}{=} \frac{1}{2}\log\left(\exp(\log 2\pi e(|a_{22}|^2\alpha P_2 + 1)) - \exp\left(\log 2\pi e\left(\frac{|a_{12}|^2}{|a_{12}|^2+|a_{22}|^2}\right)\right)\right) - \frac{1}{2}\log|a_{22}|^2\\
&\quad = \frac{1}{2}\log\left(2\pi e\left(\alpha P_2 + \frac{1}{|a_{12}|^2+|a_{22}|^2}\right)\right)
\end{aligned} \tag{125}$$

where (a) is due to the entropy power inequality, (b) holds because $a_{22}\hat{Z}_1 + \hat{Z}_1^*$ is a Gaussian random variable with zero mean and unit variance (the same as Z_2), and (c) is given by (123). Next let Z_2^* be a

Gaussian random variable, independent of all other variables, with zero mean and a variance equal to $\frac{|a_{22}|^2}{|a_{12}|^2} - 1$. We have:

$$
\begin{aligned}
H(Y_1|U,X_1,Q) &= H(a_{12}X_2 + Z_1|U,X_1,Q) \\
&= H\left(a_{22}X_2 + \tfrac{a_{22}}{a_{12}}Z_1 \middle| U,X_1,Q\right) - \frac{1}{2}\log\frac{|a_{22}|^2}{|a_{12}|^2} \\
&\overset{a}{=} H(a_{22}X_2 + Z_2 + Z_2^*|U,X_1,Q) - \frac{1}{2}\log\frac{|a_{22}|^2}{|a_{12}|^2} \\
&\overset{b}{\geq} \frac{1}{2}\log\Big(\exp\big(2H(a_{22}X_2 + Z_2|U,X_1,Q)\big) + \exp\big(2H(Z_2^*|U,X_1,Q)\big)\Big) - \frac{1}{2}\log\frac{|a_{22}|^2}{|a_{12}|^2} \\
&\overset{c}{=} \frac{1}{2}\log\left(\begin{array}{c}\exp(\log 2\pi e(|a_{22}|^2\alpha P_2 + 1)) \\ + \exp\left(\log 2\pi e\left(\frac{|a_{22}|^2}{|a_{12}|^2} - 1\right)\right)\end{array}\right) - \frac{1}{2}\log\frac{|a_{22}|^2}{|a_{12}|^2} \\
&= \frac{1}{2}\log(2\pi e(|a_{12}|^2\alpha P_2 + 1))
\end{aligned}
\tag{126}
$$

where (a) holds because $Z_2 + Z_2^*$ is a Gaussian random variable with zero mean and a variance equal to $\frac{|a_{22}|^2}{|a_{12}|^2}$, i.e., the same as $\frac{a_{22}}{a_{12}}Z_1$, (b) is due to the entropy power inequality, and (c) is given by (123). Thus, we applied the entropy power inequality twice: once to establish an upper bound on $H(\hat{Y}_1|U,X_1,Q)$ as in (125) and once to establish a lower bound on $H(Y_1|U,X_1,Q)$ as in (126). It is important to note that one may also apply the principle of "Gaussian maximizes differential entropy" to obtain an upper bound on $H(\hat{Y}_1|U,X_1,Q)$, however, the upper bound derived by that approach does not necessarily relate to α (which is specifically determined by (123)) and thereby we cannot establish a bound consistent to other entropy functions including the auxiliary random variable U.

Let again review the derivations in (126). One can easily see that all of the relations given in (126) hold only for the channel with weak interference at the first receiver where $|a_{12}| < |a_{22}|$. Next, we consider the Gaussian IC with strong interference at the first receiver where

$$
|a_{12}| \geq |a_{22}| \tag{127}
$$

For this case, the derivations in (126) are no longer valid. The fact is that when $|a_{12}| \geq |a_{22}|$, by fixing $H(Y_2|U, X_1, Q)$ as in (123), we cannot establish a lower bound on $H(Y_1|U, X_1, Q)$ using the entropy power inequality because given X_1, the output Y_1 is no longer a stochastically degraded version of Y_2. Therefore, we need to change our strategy for the optimization. For this purpose, first note that the strong interference condition (127) implies the following inequality:

$$I(X_2; Y_2|U, X_1, Q) \leq I(X_2; Y_1|U, X_1, Q) \quad \text{for all PDFs} \quad P_{QUX_1X_2} \tag{128}$$

Considering (128), we can derive:

$$\begin{aligned} &I(X_2; Y_2|U, X_1, Q) + I(U, X_1; Y_1|Q) \\ &\qquad \leq I(X_2; Y_1|U, X_1, Q) + I(U, X_1; Y_1|Q) = I(X_1, X_2; Y_1|Q) \end{aligned} \tag{129}$$

$$\begin{aligned} &I(X_2; Y_2|U, X_1, Q) + I(U, X_1; Y_1|G_1, Q) \\ &\qquad \leq I(X_2; Y_1|U, X_1, Q) + I(U, X_1; Y_1|G_1, Q) = I(X_1, X_2; Y_1|G_1, Q) \end{aligned} \tag{130}$$

Then, we evaluate $H(\hat{Y}_1|U, X_1, Q)$. We can write:

$$\begin{aligned} \frac{1}{2}\log\left(2\pi e\left(\frac{1}{|a_{12}|^2 + |a_{22}|^2}\right)\right) &= H(\hat{Y}_1|U, X_1, X_2, Q) \\ &\leq H(\hat{Y}_1|U, X_1, Q) \\ &= H\left(\frac{a_{12}Y_1 + a_{22}Y_2}{|a_{12}|^2 + |a_{22}|^2}\middle|U, X_1, Q\right) \\ &= H(X_2 + \hat{Z}_1|U, X_1, Q) \\ &\leq H(X_2 + \hat{Z}_1|Q) \leq \frac{1}{2}\log\left(2\pi e\left(P_2 + \frac{1}{|a_{12}|^2 + |a_{22}|^2}\right)\right) \end{aligned} \tag{131}$$

Comparing the two sides of (131), we can argue that there is $\alpha \in [0,1]$ such that:

$$H(\hat{Y}_1|U, X_1, Q) = \frac{1}{2}\log\left(2\pi e\left(\alpha P_2 + \frac{1}{|a_{12}|^2 + |a_{22}|^2}\right)\right) \tag{132}$$

Now by substituting (129-130) in (99) and (113), and considering (115) and (120), one can readily verify that the optimization is reduced to minimize $H(Y_1|U,X_1,Q)$ and $H(Y_2|U,X_1,Q)$, simultaneously. Moreover, given X_1, both Y_1 and Y_2 are stochastically degraded versions of $\hat{Y}_1$. Therefore, considering (132), one can successfully apply the entropy power inequality to establish lower bounds on $H(Y_1|U,X_1,Q)$ and $H(Y_2|U,X_1,Q)$. Clearly, let $\hat{Z}_1^{\nabla}$ and $\hat{Z}_1^{\Delta}$ be two Gaussian random variables, independent of all other variables, with zero mean and variances $\frac{|a_{22}|^2}{|a_{12}|^2(|a_{12}|^2+|a_{22}|^2)}$ and we have $\frac{|a_{12}|^2}{|a_{22}|^2(|a_{12}|^2+|a_{22}|^2)}$, respectively. We have:

$$
\begin{aligned}
&H(Y_1|U,X_1,Q)\\
&\quad = H(a_{12}X_2+Z_1|U,X_1,Q)\\
&\quad = H\left(X_2+\frac{1}{a_{12}}Z_1\middle|U,X_1,Q\right)-\frac{1}{2}\log\frac{1}{|a_{12}|^2}\\
&\quad \stackrel{a}{=} H\left(X_2+\hat{Z}_1+\hat{Z}_1^{\nabla}\middle|U,X_1,Q\right)-\frac{1}{2}\log\frac{1}{|a_{12}|^2}\\
&\quad \stackrel{b}{\geq} \frac{1}{2}\log\left(\exp\left(2H\left(X_2+\hat{Z}_1\middle|U,X_1,Q\right)\right)+\exp\left(2H\left(\hat{Z}_1^{\nabla}\middle|U,X_1,Q\right)\right)\right)-\frac{1}{2}\log\frac{1}{|a_{12}|^2}\\
&\quad \stackrel{c}{=} \frac{1}{2}\log\left(\begin{array}{c}\exp\left(\log 2\pi e\left(\alpha P_2+\frac{1}{|a_{12}|^2+|a_{22}|^2}\right)\right)\\ +\exp\left(\log 2\pi e\left(\frac{|a_{22}|^2}{|a_{12}|^2(|a_{12}|^2+|a_{22}|^2)}\right)\right)\end{array}\right)-\frac{1}{2}\log\frac{1}{|a_{12}|^2}\\
&\quad = \frac{1}{2}\log(2\pi e(|a_{12}|^2\alpha P_2+1))
\end{aligned}
\tag{133}
$$

where (a) holds because $\hat{Z}_1+\hat{Z}_1^{\nabla}$ is a Gaussian variable with zero mean and variance $\frac{1}{|a_{12}|^2}$, i.e., the same as $\frac{1}{a_{12}}Z_1$, (b) is due to the entropy power inequality, and (c) is given by (132). Similarly, we can derive:

$$
\begin{aligned}
&H(Y_2|U,X_1,Q)\\
&\quad = H(a_{22}X_2+Z_2|U,X_1,Q)
\end{aligned}
$$

$$= H\left(X_2 + \tfrac{1}{a_{22}} Z_2 \middle| U, X_1, Q\right) - \frac{1}{2}\log\frac{1}{|a_{22}|^2}$$

$$\overset{a}{=} H\left(X_2 + \hat{Z}_1 + \hat{Z}_1^{\Delta} \middle| U, X_1, Q\right) - \frac{1}{2}\log\frac{1}{|a_{22}|^2}$$

$$\overset{b}{\geq} \frac{1}{2}\log\left(\exp\left(2H\left(X_2 + \hat{Z}_1 \middle| U, X_1, Q\right)\right) + \exp\left(2H\left(\hat{Z}_1^{\Delta} \middle| U, X_1, Q\right)\right)\right) - \frac{1}{2}\log\frac{1}{|a_{22}|^2}$$

$$\overset{c}{=} \frac{1}{2}\log\left(\begin{array}{c}\exp\left(\log 2\pi e\left(\alpha P_2 + \dfrac{1}{|a_{12}|^2 + |a_{22}|^2}\right)\right) \\ + \exp\left(\log 2\pi e\left(\dfrac{|a_{12}|^2}{|a_{22}|^2(|a_{12}|^2 + |a_{22}|^2)}\right)\right)\end{array}\right) - \frac{1}{2}\log\frac{1}{|a_{22}|^2}$$

$$= \frac{1}{2}\log(2\pi e(|a_{22}|^2 \alpha P_2 + 1))$$

(134)

where (a) holds because $\hat{Z}_1 + \hat{Z}_1^{\Delta}$ is a Gaussian variable with zero mean and variance $\frac{1}{|a_{22}|^2}$, i.e., the same as $\frac{1}{a_{22}} Z_2$, (b) is due to the entropy power inequality, and (c) is given by (132). Therefore, $H(Y_1|U, X_1, Q)$ and $H(Y_2|U, X_1, Q)$ are minimized by the right side of (133) and (134), respectively. The proof is thus complete. ■

As indicated earlier, an outer bound was also established by Wang and Tse in [20] for the Gaussian IC with conferencing decoders. Indeed, by a straightforward comparison via simple algebraic computations, one can verify that each of the constraints given in (114) is tighter than a corresponding one of [20, Page 2920]. Thus, for all channel parameters, the outer bound $\mathfrak{R}_o^{GIC \to CD}$ given by (114) is strictly tighter than that of [20, Lemma 5.1].

We conclude this chapter by providing some numerical results. In Fig. 16, 17, and 18, we compare our new outer bound for the Gaussian IC with conferencing decoders and that of [20] in the weak, strong, and mixed interference regimes, respectively.

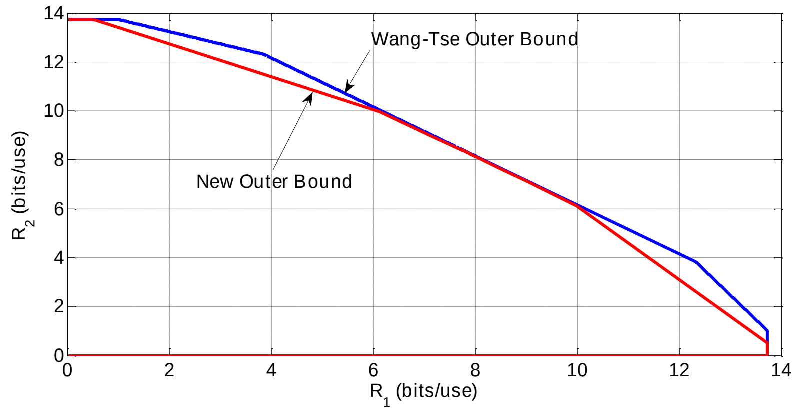

Figure 16. Comparison of the new outer bound for the Gaussian IC (9) with conferencing decoders and that of [20] in the weak interference regime ($P_1 = P_2 = 1,\ D_{12} = D_{21} = .5,\ a_{11} = a_{22} = 100,\ a_{12} = a_{21} = 60$).

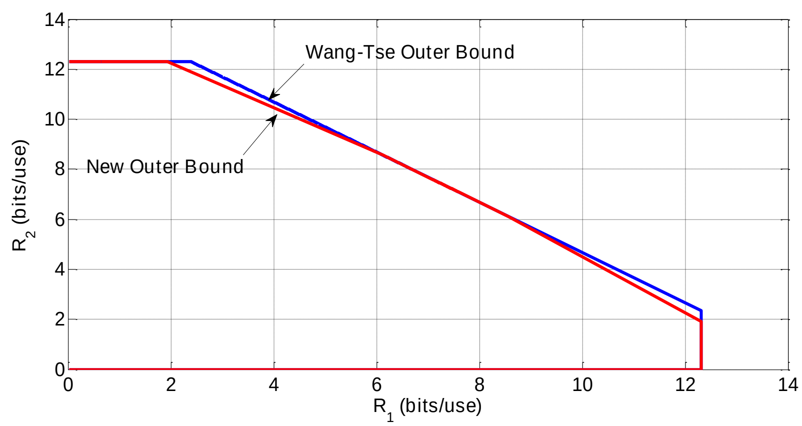

Figure 17. Comparison of the new outer bound for the Gaussian IC (9) with conferencing decoders and that of [20] in the strong interference regime ($P_1 = P_2 = 1,\ D_{12} = D_{21} = .5,\ a_{11} = a_{22} = 60,\ a_{12} = a_{21} = 100$).

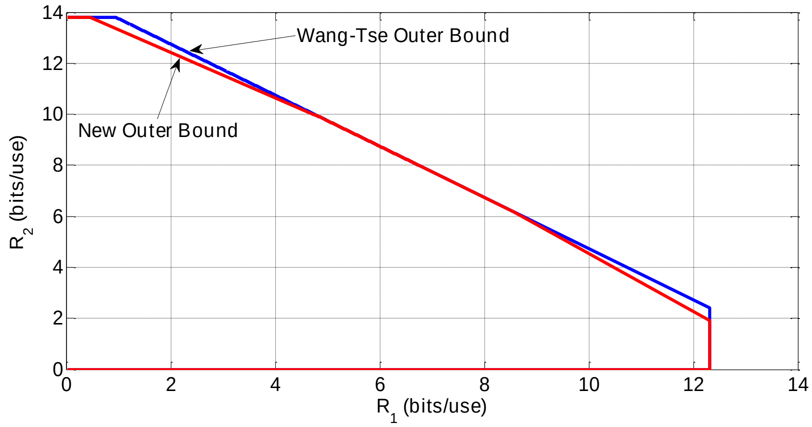

Figure 18. Comparison of the new outer bound for the Gaussian IC (9) with conferencing decoders and that of [20] in the mixed interference regime ($P_1 = P_2 = 1,\ D_{12} = D_{21} = .5,\ a_{11} = a_{21} = 60,\ a_{12} = a_{22} = 100$).

As shown in these figures, for all cases our new outer bound is strictly tighter.

Conclusion

In this thesis, we established a full characterization of the sum-rate capacity for degraded interference networks with any number of transmitters, any number of receivers, and any possible distribution of messages among transmitters and receivers. We proved that the sum-rate capacity of these networks can be achieved by a successive decoding scheme. Moreover, using graphical algorithms, we clearly identified those messages which should be considered in the transmission scheme to achieve the sum-rate capacity. Next, we derived a unified outer bound on the sum-rate capacity of general non-degraded interference networks. To this end, our idea was to enhance the non-degraded network with artificial outputs to obtain a degraded network whose capacity region includes that of the original network as a subset. Thus, the sum-rate capacity expression of the artificial degraded network would be an outer bound on the sum-rate capacity of the original network as well. We proved that the derived outer bound is sum-rate optimal for several variations of degraded networks. In particular, we obtained sum-rate capacities for interesting scenarios such as generalized Z-interference networks and many-to-one interference networks. Also, for the first time, we identified networks for which the incorporation of both successive decoding and treating interference as noise achieves their sum-rate capacity.

Next, by taking insight from the results for degraded networks, we presented an extension to the standard cut-set bound for general communication networks which we refer to as nested cut-set bound. To derive this bound, we apply a series of cuts in a nested configuration to the network first and then bound the information rate that flows through the cuts. The key idea for bounding step is to impose a degraded arrangement among the receivers corresponding to the cuts. Therefore, the bound is in fact a generalization of the outer bound for interference networks: here cooperative relaying nodes are introduced into the problem as well but the proof style for the derivation of the outer bound remains the same. The nested cut-set bound, which uniformly holds for all general communication networks of arbitrary large sizes where any subset of nodes may cooperatively communicate to any other subset of them, is indeed tighter than the cut-set bound for networks with more than one receiver.

Finally, we investigated capacity bounds for the two-user interference channel with cooperative receivers via conferencing links of finite capacities. By applying new techniques, we presented novel capacity outer bounds for this channel. Using the outer bounds, we proved several new capacity results for interesting channels with unidirectional cooperation in strong and mixed interference regimes. A fact is that a conferencing link (between receivers) may be utilized to provide one receiver with information about its corresponding signal or its non-corresponding signal (interference). An interesting conclusion of our work was to show that both of these strategies can be helpful to achieve the capacity of the channel. Lastly, for the case of Gaussian IC with conferencing receivers, we argued that our outer bound is strictly tighter than the previous one derived by Wang and Tse.

References

[1] R. K. Farsani, "Fundamental limits of communications in interference networks-Part I: Basic structures," *2012, available at* http://arxiv.org/abs/1207.3018.

[2] ___________, "On the capacity region of the broadcast, the interference, and the cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2600–2623, May 2015.

[3] R. K. Farsani and F. Marvasti, "Interference networks with general message sets: A random coding scheme", 49th *Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2011.

[4] R. K. Farsani, "How much rate splitting is required for a random coding scheme? A new achievable rate region for the broadcast channel with cognitive relays," 50th *Annual Allerton Conference on Communication, Control, and Computing*., Monticello, IL, Oct. 2012.

[5] J. Jose and S. Vishwanath, "Sum capacity of k-user Gaussian degraded interference channels," *submitted to IEEE Transactions on Information Theory*. arXiv:1004.2104.

[6] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[7] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, 2008.

[8] A. El Gamal, and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[9] I. Shomorony and A. S. Avestimehr, "A generalized cut-set bound for deterministic multi-flow networks and its applications," in *IEEE International Symposium on Information Theory Proceedings*, 2014, pp. 271-275.

[10] R. K. Farsani, "Fundamental limits of communications in interference networks-Part II: Information flow in degraded networks," 2012, *available at* http://arxiv.org/abs/1207.3027.

[11] ___________, "The sum-rate capcity of general degarded interference networks with arbitrary topologies," in *IEEE International Symposium on Information Theory Proceedings*, 2014, pp. 276-280.

[12] G. Kramer, I. Maric, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, no. 3, pp. 271–425, 2006.

[13] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[14] F. M. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 441–445, 1983.

[15] A. Haghi, R. Khosravi-Farsani, M. R. Aref, and F. Marvasti, “The capacity region of p-transmitter/q-receiver multiple-access channels with common information,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7359–7376, 2011.

[16] I. Maric, R. D. Yates, and G. Kramer, “Capacity of interference channels with partial transmitter cooperation,” *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3536–3548, 2007.

[17] R. Dabora and S. D. Servetto, “Broadcast channels with cooperating decoders,” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5438–5454, 2006.

[18] O. Simeone, D. Gunduz, H. V. Poor, A. J. Goldsmith, and S. Shamai, “Compound multiple-access channels with partial cooperation,” *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2425–2441, 2009.

[19] I-H. Wang and D. N. Tse, “Interference mitigation through limited transmitter cooperation,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2941–2965, 2011.

[20] ______________________, “Interference mitigation through limited receiver cooperation,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2913–2940, 2011.

[21] A. Host-Madsen, “Capacity bounds for cooperative diversity,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 15221544, Apr. 2006.

[22] H. T. Do, T. J. Oechtering, and M. Skoglund, “An achievable rate region for the gaussian z-interference channel with conferencing,” in *47th Annual Allerton Conference on Communication, Control, and Computing*, 2009, pp. 75–81.

[23] H. Bagheri, A. S. Motahari, and A. K. Khandani, “On the symmetric gaussian interference channel with partial unidirectional cooperation,” arXiv preprint arXiv:0909.2777, 2009.

[24] Y. Cao, and B. Chen, “An achievable rate region for interference channels with conferencing,” in *IEEE International Symposium on Information Theory*, 2007, pp. 1251–1255.

[25] L. Zhou, and W. Yu, “Gaussian z-interference channel with a relay link: Achievability region and asymptotic sum capacity,” *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2413–2426, 2012.

[26] N. Liu, D. Gunduz, and W. Kang, “Capacity results for a class of deterministic z-interference channels with unidirectional receiver conferencing,” in *International ICST Conference on Communications and Networking in China*, Aug 2011, pp. 580–584.

[27] M. H. Costa, and A. El Gamal, “The capacity region of the discrete memoryless interference channel with strong interference.” *IEEE Transactions on Information Theory*, vol. 33, no. 5, pp. 710–711, 1987.

[28] H. Sato, “Two-user communication channels,” *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 295–304, 1977.

[29] D. Tuninetti, “An outer bound for the memoryless two-user interference channel with general cooperation,” in *IEEE Information Theory Workshop*, 2012, pp. 217–221.

[30] I. Csiszar, and J. Korner, “Broadcast channels with confidential messages,” *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[31] A. Jovicic, H.Wang, and P. Viswanath, “On network interference management,” in *IEEE Trans. Inform. Theory*, vol. 56, pp. 4941– 4955, Oct. 2010.

[32] G. Bresler, A. Parekh, and D. Tse, “The approximate capacity of the many-to-one and one-to-many Gaussian interference channels,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566-4592, Sep. 2010.

[33] V. S. Annapureddy, and V. V. Veeravalli, “Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region,” in *IEEE Trans. Inform. Theory*, vol. 55(7), July 2009, pp. 3032 – 3050.

[34] R. K. Farsani, “The k-user interference channel: Strong interference regime,” in *IEEE International Symposium on Information Theory Proceedings*, 2013, pp. 2029–2033.

[35] ____________, “On the capacity region of the two-user interference channel,” in *IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 2734–2738.

Appendix A

Converse Proof of Proposition 3.1

We show that the Gaussian distributions are optimal for the sum-rate capacity expression given in (45). There are several ways to prove this result. Here, we make use of the entropy power inequality. Consider the argument of the maximization (45). Fix a joint PDF $P_Q P_{X_1 X_2|Q} P_{X_3|Q} P_{X_4|X_1 X_2 X_3 Q}$. We have:

$$
\begin{aligned}
&I(X_4;Y_1|X_1,X_2,X_3,Q) + I(X_3;Y_2|X_1,X_2,Q) + I(X_1,X_2;Y_3|Q) \\
&\quad = H(Y_1|X_1,X_2,X_3,Q) - H(Y_2|X_1,X_2,X_3,Q) + H(Y_2|X_1,X_2,Q) - H(Y_3|X_1,X_2,Q) \\
&\qquad + H(Y_3|Q) - H(Y_1|X_1,X_2,X_3,X_4,Q) \\
&\quad = H(a_4X_4 + Z_1|X_1,X_2,X_3,Q) - H\left(\tfrac{1}{b_2}(a_4X_4 + Z_1) + \sqrt{1-\tfrac{1}{b_2^2}}\,\tilde{Z}_2 \middle| X_1,X_2,X_3,Q\right) \\
&\qquad + H\left(\tfrac{a_3}{b_2}X_3 + \tfrac{a_4}{b_2}X_4 + Z_2 \middle| X_1,X_2,Q\right) - H\left(\tfrac{1}{b_3}\left(\tfrac{a_3}{b_2}X_3 + \tfrac{a_4}{b_2}X_4 + Z_2\right) + \sqrt{1-\tfrac{1}{b_3^2}}\,\tilde{Z}_3 \middle| X_1,X_2,Q\right) \\
&\qquad + H\left(\tfrac{a_1}{b_2b_3}X_1 + \tfrac{a_2}{b_2b_3}X_2 + \tfrac{a_3}{b_2b_3}X_3 + \tfrac{a_4}{b_2b_3}X_4 + Z_3 \middle| Q\right) - H(Z_1)
\end{aligned}
\tag{A~1}
$$

Next let $X_1^G, X_2^G, X_3^G, X_4^G$ be jointly Gaussian RVs with a covariance matrix identical to that of X_1, X_2, X_3, X_4. Thus, we can decompose X_4^G as follows:

$$
X_4^G = \alpha\sqrt{\frac{P_4}{P_1}}X_1^G + \beta\sqrt{\frac{P_4}{P_3}}X_3^G + \sqrt{\left(1-(\alpha^2+\beta^2)\right)P_4}\,Z
\tag{A~2}
$$

where α, β belong to the interval $[-1,1]$ with $\alpha^2 + \beta^2 \leq 1$, and Z is a Gaussian RV independent of X_1^G, X_2^G, X_3^G with zero mean and unit variance. Consider the expressions on the right hand side of (A~1). For the first term, we have:

$$
\begin{aligned}
\frac{1}{2}\log(2\pi e) &\leq H(Z_1) \leq H(a_4X_4 + Z_1|X_1,X_2,X_3,Q) \leq H(a_4X_4 + Z_1|X_1,X_3) \\
&\leq H\left(a_4X_4^G + Z_1 \middle| X_1^G, X_3^G\right) = \tfrac{1}{2}\log\left(2\pi e\left(a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4 + 1\right)\right)
\end{aligned}
\tag{A~3}
$$

Comparing the two sides of (A~3), we deduce that there exists a $\lambda_1 \in [0,1]$ such that:

$$H(a_4X_4 + Z_1|X_1, X_2, X_3, Q) = \frac{1}{2}\log\left(2\pi e\left(\lambda_1 a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4+1\right)\right) \tag{A~4}$$

Then consider the second term in (A~1). We can write:

$$\begin{aligned}
&H\left(\frac{1}{b_2}(a_4X_4 + Z_1) + \sqrt{1-\frac{1}{b_2^2}}\,\tilde{Z}_2\middle|X_1, X_2, X_3, Q\right)\\
&\qquad\overset{(a)}{\geq} \frac{1}{2}\log\left(2^{2H\left(\frac{1}{b_2}(a_4X_4+Z_1)\middle|X_1,X_2,X_3,Q\right)} + 2^{2H\left(\sqrt{1-\frac{1}{b_2^2}}\,\tilde{Z}_2\middle|X_1,X_2,X_3,Q\right)}\right)\\
&\qquad\overset{(b)}{=} \frac{1}{2}\log\left(2\pi e\frac{1}{b_2^2}\left(\lambda_1 a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4+1\right) + 2\pi e\left(1-\frac{1}{b_2^2}\right)\right)\\
&\qquad= \frac{1}{2}\log\left(2\pi e\left(\lambda_1\frac{a_4^2}{b_2^2}\left(1-(\alpha^2+\beta^2)\right)P_4+1\right)\right)
\end{aligned} \tag{A~5}$$

where (a) is due to the entropy power inequality and (b) is due to (A~4). Thereby, from (A~4) and (A~5) we obtain:

$$\begin{aligned}
&H(a_4X_4 + Z_1|X_1, X_2, X_3, Q) - H\left(\frac{1}{b_2}(a_4X_4 + Z_1) + \sqrt{1-\frac{1}{b_2^2}}\,\tilde{Z}_2\middle|X_1, X_2, X_3, Q\right)\\
&\qquad\leq \frac{1}{2}\log\left(\frac{\lambda_1 a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4+1}{\lambda_1\frac{a_4^2}{b_2^2}\left(1-(\alpha^2+\beta^2)\right)P_4+1}\right) \overset{(a)}{\leq} \frac{1}{2}\log\left(\frac{a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4+1}{\frac{a_4^2}{b_2^2}\left(1-(\alpha^2+\beta^2)\right)P_4+1}\right)
\end{aligned} \tag{A~6}$$

where inequality (a) holds because the expression on its left hand side is monotonically increasing in terms of λ_1 (for the case where $b_2 \geq 1$).

Next we evaluate the third and the forth terms in (A~1). We have:

$$\begin{aligned}
\frac{1}{2}\log(2\pi e) &\leq H(Z_2) \leq H\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2\middle|X_1, X_2, Q\right)\\
&\leq H\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2\middle|X_1\right) \leq H\left(\frac{a_3}{b_2}X_3^G + \frac{a_4}{b_2}X_4^G + Z_2\middle|X_1^G\right)
\end{aligned}$$

$$\leq \frac{1}{2}\log\left(2\pi e\left(\frac{1}{b_2^2}\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + \frac{a_4^2}{b_2^2}\left(1-(\alpha^2+\beta^2)\right)P_4 + 1\right)\right) \tag{A~7}$$

Comparing the two sides of (A~7), we deduce that there exists a $\lambda_2 \in [0,1]$ such that:

$$H\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2 \middle| X_1, X_2, Q\right) = \frac{1}{2}\log\left(2\pi e\left(\lambda_2 \frac{1}{b_2^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1\right)\right) \tag{A~8}$$

Considering (A~8) and using the entropy power inequality, one can derive:

$$H\left(\frac{1}{b_3}\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2\right) + \sqrt{1-\frac{1}{b_3^2}}\ \tilde{Z}_3 \middle| X_1, X_2, Q\right) \geq \frac{1}{2}\log\left(2\pi e\left(\lambda_2 \frac{1}{b_2^2 b_3^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1\right)\right) \tag{A~9}$$

Now from (A~8) and (A~9), we obtain:

$$\begin{aligned}
&H\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2 \middle| X_1, X_2, Q\right) - H\left(\frac{1}{b_3}\left(\frac{a_3}{b_2}X_3 + \frac{a_4}{b_2}X_4 + Z_2\right) + \sqrt{1-\frac{1}{b_3^2}}\ \tilde{Z}_3 \middle| X_1, X_2, Q\right) \\
&\leq \frac{1}{2}\log\left(\frac{\lambda_2 \frac{1}{b_2^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1}{\lambda_2 \frac{1}{b_2^2 b_3^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1}\right) \\
&\overset{(a)}{\leq} \frac{1}{2}\log\left(\frac{\frac{1}{b_2^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1}{\frac{1}{b_2^2 b_3^2}\left(\left(|a_3| + |a_4\beta|\sqrt{\frac{P_4}{P_3}}\right)^2 + a_4^2\left(1-(\alpha^2+\beta^2)\right)P_4\right) + 1}\right)
\end{aligned} \tag{A~10}$$

where inequality (a) holds because the expression on its left hand side is monotonically increasing in terms of λ_2 (for the case where $b_2, b_3 \geq 1$). Finally, consider the fifth term of (A~1). We have:

$$
\begin{aligned}
&H\left(\frac{a_1}{b_2 b_3}X_1 + \frac{a_2}{b_2 b_3}X_2 + \frac{a_3}{b_2 b_3}X_3 + \frac{a_4}{b_2 b_3}X_4 + Z_3 \middle| Q\right) \\
&\qquad \leq H\left(\frac{a_1}{b_2 b_3}X_1^G + \frac{a_2}{b_2 b_3}X_2^G + \frac{a_3}{b_2 b_3}X_3^G + \frac{a_4}{b_2 b_3}X_4^G + Z_3\right) \\
&\qquad \leq \frac{1}{2}\log\left(2\pi e\left(\frac{1}{b_2^2 b_3^2}\begin{pmatrix} a_1^2 P_1 + a_2^2 P_2 + a_3^2 P_3 + a_4^2 P_4 \\ +2\left|a_1 a_2 \mathbb{E}\left[X_1^G X_2^G\right]\right| + 2\left|a_1 a_4 \mathbb{E}\left[X_1^G X_4^G\right]\right| \\ +2\left|a_2 a_4 \mathbb{E}\left[X_2^G X_4^G\right]\right| + 2\left|a_3 a_4 \mathbb{E}\left[X_3^G X_4^G\right]\right| \end{pmatrix} + 1\right)\right) \\
&\qquad \overset{(a)}{=} \frac{1}{2}\log\left(2\pi e\left(\frac{1}{b_2^2 b_3^2}\begin{pmatrix} a_1^2 P_1 + a_2^2 P_2 + a_3^2 P_3 + a_4^2 P_4 \\ +2\left|a_1 a_2 \mathbb{E}\left[X_1^G X_2^G\right]\right| + 2|a_1 a_4 \alpha|\sqrt{P_1 P_4} + \\ 2\left|a_2 a_4 \alpha \mathbb{E}\left[X_1^G X_2^G\right]\right|\sqrt{\frac{P_4}{P_1}} + 2|a_3 a_4 \beta|\sqrt{P_3 P_4} \end{pmatrix} + 1\right)\right) \\
&\qquad \overset{(b)}{\leq} \frac{1}{2}\log\left(2\pi e\left(\frac{1}{b_2^2 b_3^2}\begin{pmatrix} a_1^2 P_1 + a_2^2 P_2 + a_3^2 P_3 + a_4^2 P_4 \\ +2|a_1 a_2|\sqrt{P_1 P_2} + 2|a_1 a_4 \alpha|\sqrt{P_1 P_4} + \\ 2|a_2 a_4 \alpha|\sqrt{P_2 P_4} + 2|a_3 a_4 \beta|\sqrt{P_3 P_4} \end{pmatrix} + 1\right)\right)
\end{aligned} \tag{A~11}
$$

where equality (a) is due to (A~2) and inequality (b) holds because $\mathbb{E}\left[X_1^G X_2^G\right] \leq \sqrt{P_1 P_2}$. By substituting (A~6), (A~10), and (A~11) in (A~1), we derive the desired result. ■

Appendix B
Converse Proof of Proposition 3.2

We prove that the Gaussian inputs are optimal for (55). Consider the argument of the maximization (55); we can write:

$$I(X_1, X_2; Y_1|W) + I(W; Y_2)$$
$$= H(X_1 + aX_2 + Z_1|W) - H\left(b(X_1 + aX_2 + Z_1) + \sqrt{1-b^2}\,\tilde{Z}_2\middle|W\right) + H(bX_1 + X_2 + Z_2) - H(Z_1)$$

(A~12)

Define α and β as follows:

$$\alpha \triangleq \mathrm{sign}(b)\sqrt{\frac{\mathbb{E}[(\mathbb{E}[X_1|W])^2]}{\mathbb{E}[X_1^2]}}, \qquad \beta \triangleq \sqrt{\frac{\mathbb{E}[(\mathbb{E}[X_2|W])^2]}{\mathbb{E}[X_2^2]}}$$

(A~13)

where $\mathrm{sign}(b)$ is equal to 1 if $b > 0$, and is equal to -1 if $b < 0$. Note that we have:

$$\mathbb{E}\left[X_i^2\right] - \mathbb{E}[(\mathbb{E}[X_i|W])^2] = \mathbb{E}\left[\mathbb{E}\left[X_i^2|W\right] - (\mathbb{E}[X_i|W])^2\right] = \mathbb{E}\left[\mathbb{E}^2[X_i|W]\right] \geq 0, \quad i = 1,2$$

(A~14)

where $\mathbb{E}^2[X_i|W]$ is derived from Definition II.2 of [1]. Therefore, α^2 and β^2 both belong to the interval $[0,1]$. Then consider the expressions on the right side of (A~12). We have:

$$H(X_1 + aX_2 + Z_1|W) \geq H(X_1 + aX_2 + Z_1|X_1, X_2, W) = \frac{1}{2}\log(2\pi e)$$

(A~15)

Also, using the "Gaussian maximizes entropy" principle, we can derive:

$$H(X_1 + aX_2 + Z_1|W) \leq \mathbb{E}\left[\frac{1}{2}\log\left(2\pi e(\mathbb{E}^2[X_1 + aX_2 + Z_1|W])\right)\right]$$
$$\overset{(a)}{=} \mathbb{E}\left[\frac{1}{2}\log\left(2\pi e(\mathbb{E}^2[X_1|W] + \mathbb{E}^2[aX_2|W] + 1)\right)\right]$$
$$\overset{(b)}{\leq} \frac{1}{2}\log\left(2\pi e\left(\mathbb{E}\left[\mathbb{E}^2[X_1|W]\right] + a^2\mathbb{E}\left[\mathbb{E}^2[X_2|W]\right] + 1\right)\right)$$

$$= \frac{1}{2}\log\left(2\pi e\begin{pmatrix} \mathbb{E}[\mathbb{E}[X_1^2|W] - (\mathbb{E}[X_1|W])^2] \\ +a^2\mathbb{E}[\mathbb{E}[X_2^2|W] - (\mathbb{E}[X_2|W])^2] + 1 \end{pmatrix}\right)$$

$$\overset{(c)}{=} \frac{1}{2}\log\left(2\pi e\left((1-\alpha^2)P_1 + a^2(1-\beta^2)P_2 + 1\right)\right)$$

(A~16)

where (a) is derived by Lemma II.1 of [1], (b) is due to Jensen inequality, and (c) is derived by (A~13). Based on (A~15) and (A~16), we deduce that there exist μ_α and μ_β with:

$$H(X_1 + aX_2 + Z_1|W) = \frac{1}{2}\log\left(2\pi e\left((1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2 + 1\right)\right),$$

$$\alpha^2 \le \mu_\alpha \le 1, \qquad \beta^2 \le \mu_\beta \le 1$$

(A~17)

Now for the second expression of (A~12), using the entropy power inequality, we have:

$$H\left(b(X_1 + aX_2 + Z_1) + \sqrt{1-b^2}\,\tilde{Z}_2\middle|W\right)$$

$$\ge \frac{1}{2}\log\left(2^{2H\left(b(X_1 + aX_2 + Z_1)\middle|W\right)} + 2^{2H\left(\sqrt{1-b^2}\,\tilde{Z}_2\middle|W\right)}\right)$$

$$= \frac{1}{2}\log\left(2\pi e\left(b^2\left((1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2\right) + 1\right)\right)$$

(A~18)

By combining (A~17) and (A~18), we obtain:

$$H(X_1 + aX_2 + Z_1|W) - H\left(b(X_1 + aX_2 + Z_1) + \sqrt{1-b^2}\,\tilde{Z}_2\middle|W\right)$$

$$\le \frac{1}{2}\log\left(\frac{(1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2 + 1}{b^2\left((1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2\right) + 1}\right)$$

(A~19)

Then, consider the following deterministic function:

$$f(\mu_\alpha, \mu_\beta) \triangleq \frac{(1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2 + 1}{b^2\left((1-\mu_\alpha)P_1 + a^2(1-\mu_\beta)P_2\right) + 1}$$

(A~20)

It is easy to show that for the case of $ab = 1, |a| \geq 1$, the function $f(\mu_\alpha, \mu_\beta)$ is monotonically decreasing in terms of both μ_α and μ_β. Thereby, since $\alpha^2 \leq \mu_\alpha$ and $\beta^2 \leq \mu_\beta$, we derive:

$$\begin{aligned}
&H(X_1 + aX_2 + Z_1|W) - H\left(b(X_1 + aX_2 + Z_1) + \sqrt{1-b^2}\,\tilde{Z}_2\middle|W\right) \\
&\qquad \leq \frac{1}{2}\log\left(\frac{(1-\alpha^2)P_1 + a^2(1-\beta^2)P_2 + 1}{b^2\left((1-\alpha^2)P_1 + a^2(1-\beta^2)P_2\right) + 1}\right)
\end{aligned} \tag{A~21}$$

Finally, for the third expression on the right side of (A~12), we have:

$$\begin{aligned}
H(bX_1 + X_2 + Z_2) &\leq \frac{1}{2}\log\left(2\pi e(b^2P_1 + P_2 + 2b\mathbb{E}[X_1X_2] + 1)\right) \\
&\overset{(a)}{\leq} \frac{1}{2}\log\left(2\pi e\left(b^2P_1 + P_2 + 2|b|\sqrt{\mathbb{E}[(\mathbb{E}[X_1|W])^2]}\sqrt{\mathbb{E}[(\mathbb{E}[X_1|W])^2]} + 1\right)\right) \\
&\overset{(b)}{=} \frac{1}{2}\log\left(2\pi e\left(b^2P_1 + P_2 + 2b\alpha\beta\sqrt{P_1P_2} + 1\right)\right)
\end{aligned} \tag{A~22}$$

where inequality (a) is due to Lemma II.1 of [1] and equality (b) is due to (A~13). Now by substituting (A~21) and (A~22) in (A~12), we obtain the desired result. ■

Appendix C

Proof of Theorem 6.1

Consider a length-n code with vanishing average probability of error. Define new auxiliary random variables:

$$U_t \triangleq \left(M_1, Y_2^{t-1}, Y_{1,t+1}^n\right)$$
$$V_t \triangleq \left(M_2, Y_1^{t-1}, Y_{2,t+1}^n\right), \quad \text{for} \quad t = 1, \dots, n$$

(A~23)

Let first derive some bounds on R_1. By Fano's inequality,

$$\begin{aligned}
nR_1 &\leq I\left(M_1; Y_1^n, V_{21}^{L_d}\right) + n\epsilon_n^1 \\
&= I(M_1; Y_1^n) + I\left(M_1; V_{21}^{L_d} \middle| Y_1^n\right) + n\epsilon_n^1 \\
&\overset{a}{\leq} I(M_1; Y_1^n | M_2) + H\left(V_{21}^{L_d}\right) + n\epsilon_n^1 \\
&\overset{b}{\leq} I(M_1; Y_1^n | M_2) + nD_{21} + n\epsilon_n^1 \\
&\leq \textstyle\sum_{t=1}^n I\left(X_{1,t}; Y_{1,t} \middle| X_{2,t}\right) + nD_{21} + n\epsilon_n^1
\end{aligned}$$

(A~24)

where inequality (a) holds because conditioning does not reduce the entropy, and inequality (b) is due to (11). Moreover,

$$\begin{aligned}
nR_1 &\leq I\left(M_1; Y_1^n, V_{21}^{L_d}\right) + n\epsilon_n^1 \\
&= I(M_1; Y_1^n) + I\left(M_1; V_{21}^{L_d} \middle| Y_1^n\right) + n\epsilon_n^1 \\
&\leq \textstyle\sum_{t=1}^n I\left(M_1; Y_{1,t} \middle| Y_{1,t+1}^n\right) + H\left(V_{21}^{L_d}\right) + n\epsilon_n^1 \\
&\overset{a}{\leq} \textstyle\sum_{t=1}^n I\left(M_1, Y_{1,t+1}^n, Y_2^{t-1}, X_{1,t}; Y_{1,t}\right) + H\left(V_{21}^{L_d}\right) + n\epsilon_n^1 \\
&\leq \textstyle\sum_{t=1}^n I\left(U_t, X_{1,t}; Y_{1,t}\right) + nD_{21} + n\epsilon_n^1
\end{aligned}$$

(A~25)

where inequality (a) holds because conditioning does not reduce the entropy. Next we derive some bounds on R_2. By Fano's inequality we have:

$$
\begin{aligned}
nR_2 &\le I\left(M_2; Y_2^n, V_{12}^{L_d}\right) + n\epsilon_n^2 \\
&\le I\left(M_2; Y_1^n, Y_2^n, V_{12}^{L_d}\right) + n\epsilon_n^2 \\
&\overset{a}{=} I(M_2; Y_1^n, Y_2^n) + n\epsilon_n^2 \\
&\le I(M_2; Y_1^n, Y_2^n | M_1) + n\epsilon_n^2
\end{aligned}
\tag{A~26}
$$

where (a) holds because $V_{12}^{L_d}$ is given by a deterministic function of (Y_1^n, Y_2^n). Now consider the mutual information function on the right side of the last inequality of (A~26). We can write:

$$
\begin{aligned}
I(M_2; Y_1^n, Y_2^n | M_1) &= I(M_2; Y_2^n | M_1) + I(M_2; Y_1^n | Y_2^n, M_1) \\
&\overset{a}{=} \sum_{t=1}^n I\left(X_{2,t}; Y_{2,t} \middle| X_{1,t}, M_1\right) \\
&\qquad + \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t} \middle| Y_{2,t}, X_{1,t}, M_1, Y_2^{t-1}, Y_{2,t+1}^n, Y_{1,t+1}^n\right) \\
&= \sum_{t=1}^n I\left(X_{2,t}; Y_{2,t} \middle| X_{1,t}, M_1\right) \\
&\qquad + \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t} \middle| Y_{2,t}, X_{1,t}, U_t, Y_{2,t+1}^n\right) \\
&\overset{b}{\le} \sum_{t=1}^n I\left(X_{2,t}; Y_{2,t} \middle| X_{1,t}\right) + \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t} \middle| Y_{2,t}, X_{1,t}, U_t\right)
\end{aligned}
\tag{A~27}
$$

where (a) holds because $X_{i,t}$ is given by a deterministic function of M_i; the inequality (b) holds because conditioning does not reduce the entropy and also given the inputs $X_{1,t}, X_{2,t}$, the outputs $Y_{1,t}, Y_{2,t}$ are independent of other variables. Similarly, we can derive:

$$
\begin{aligned}
I(M_2; Y_1^n, Y_2^n | M_1) &= I(M_2; Y_1^n | M_1) + I(M_2; Y_2^n | Y_1^n, M_1) \\
&\le \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t} \middle| X_{1,t}\right) + \sum_{t=1}^n I\left(X_{2,t}; Y_{2,t} \middle| Y_{1,t}, X_{1,t}, U_t\right)
\end{aligned}
\tag{A~28}
$$

Now by substituting (A~27) and (A~28) in (A~26), we obtain:

$$nR_2 \leq \min \left\{ \begin{array}{l} \sum_{t=1}^{n} I\left(X_{2,t}; Y_{2,t} \middle| X_{1,t}\right) + I\left(X_{2,t}; Y_{1,t} \middle| Y_{2,t}, X_{1,t}, U_t\right), \\ \sum_{t=1}^{n} I\left(X_{2,t}; Y_{1,t} \middle| X_{1,t}\right) + I\left(X_{2,t}; Y_{2,t} \middle| Y_{1,t}, X_{1,t}, U_t\right) \end{array} \right\}$$

(A~29)

Finally, we establish constraints on the sum-rate. Based on Fano's inequality, one can write:

$$\begin{aligned} n(R_1 + R_2) &\leq I\left(M_1; Y_1^n, V_{21}^{L_d}\right) + I\left(M_2; Y_2^n, V_{12}^{L_d}\right) + n(\epsilon_n^1 + \epsilon_n^2) \\ &\leq I(M_1; Y_1^n) + I(M_2; Y_2^n) + I\left(M_1; V_{21}^{L_d} \middle| Y_1^n\right) + I\left(M_2; V_{12}^{L_d} \middle| Y_2^n\right) + n(\epsilon_n^1 + \epsilon_n^2) \\ &\leq I(M_1; Y_1^n) + I(M_2; Y_2^n | M_1) + H\left(V_{21}^{L_d}\right) + H\left(V_{12}^{L_d}\right) + n(\epsilon_n^1 + \epsilon_n^2) \\ &\leq I(M_1; Y_1^n) + I(M_2; Y_2^n | M_1) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \end{aligned}$$

(A~30)

The sum of the two mutual information functions on the right side of (A~30) can be bounded as follows:

$$\begin{aligned} &I(M_1; Y_1^n) + I(M_2; Y_2^n | M_1) \\ &\quad = \sum_{t=1}^{n} I\left(M_1; Y_{1,t} \middle| Y_{1,t+1}^n\right) + \sum_{t=1}^{n} I\left(M_2; Y_{2,t} \middle| M_1, Y_2^{t-1}\right) \\ &\quad \leq \sum_{t=1}^{n} I\left(Y_2^{t-1}, M_1; Y_{1,t} \middle| Y_{1,t+1}^n\right) - \sum_{t=1}^{n} I\left(Y_2^{t-1}; Y_{1,t} \middle| M_1, Y_{1,t+1}^n\right) \\ &\qquad\qquad + \sum_{t=1}^{n} I\left(Y_{1,t+1}^n, M_2; Y_{2,t} \middle| M_1, Y_2^{t-1}\right) \\ &\quad = \sum_{t=1}^{n} I\left(Y_2^{t-1}, M_1; Y_{1,t} \middle| Y_{1,t+1}^n\right) - \sum_{t=1}^{n} I\left(Y_2^{t-1}; Y_{1,t} \middle| M_1, Y_{1,t+1}^n\right) \\ &\qquad\qquad + \sum_{t=1}^{n} I\left(Y_{1,t+1}^n; Y_{2,t} \middle| M_1, Y_2^{t-1}\right) \\ &\qquad\qquad + \sum_{t=1}^{n} I\left(M_2; Y_{2,t} \middle| M_1, Y_2^{t-1}, Y_{1,t+1}^n\right) \\ &\quad \stackrel{a}{=} \sum_{t=1}^{n} I\left(Y_2^{t-1}, M_1; Y_{1,t} \middle| Y_{1,t+1}^n\right) + \sum_{t=1}^{n} I\left(M_2; Y_{2,t} \middle| M_1, Y_2^{t-1}, Y_{1,t+1}^n\right) \\ &\quad \stackrel{b}{\leq} \sum_{t=1}^{n} I\left(Y_2^{t-1}, Y_{1,t+1}^n, M_1; Y_{1,t}\right) + \sum_{t=1}^{n} I\left(M_2; Y_{2,t} \middle| M_1, Y_2^{t-1}, Y_{1,t+1}^n\right) \\ &\quad = \sum_{t=1}^{n} I\left(U_t, X_{1,t}; Y_{1,t}\right) + \sum_{t=1}^{n} I\left(X_{2,t}; Y_{2,t} \middle| U_t, X_{1,t}\right) \end{aligned}$$

(A~31)

where (a) holds because due to the Csiszar-Korner identity the second and the third mutual information functions on the left side of (a) are equal; (b) holds because conditioning does not reduce the entropy. Then by substituting (A~31) in (A~30), we derive:

$$n(R_1+R_2) \leq \sum_{t=1}^{n} I\big(U_t, X_{1,t}; Y_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t}\big|U_t, X_{1,t}\big) + n(D_{12}+D_{21}) + n(\epsilon_n^1+\epsilon_n^2) \tag{A~32}$$

Also, we have:

$$\begin{aligned}
n(R_1+R_2) &\leq I\big(M_1; Y_1^n, V_{21}^{L_d}\big) + I\big(M_2; Y_2^n, V_{12}^{L_d}\big) + n(\epsilon_n^1+\epsilon_n^2)\\
&\leq I(M_1; Y_1^n) + I\big(M_2; Y_1^n, Y_2^n, V_{12}^{L_d}\big) + I\big(M_1; V_{21}^{L_d}\big|Y_1^n\big) + n(\epsilon_n^1+\epsilon_n^2)\\
&\overset{a}{\leq} I(M_1; Y_1^n) + I(M_2; Y_1^n, Y_2^n) + H\big(V_{21}^{L_d}\big) + n(\epsilon_n^1+\epsilon_n^2)\\
&\leq I(M_1; Y_1^n) + I(M_2; Y_1^n, Y_2^n|M_1) + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)\\
&= I(M_1; Y_1^n) + I(M_2; Y_2^n|M_1) + I(M_2; Y_1^n|Y_2^n, M_1) + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)\\
&\overset{b}{\leq} \sum_{t=1}^{n} I\big(U_t, X_{1,t}; Y_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t}\big|U_t, X_{1,t}\big)\\
&\qquad + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{1,t}\big|Y_{2,t}, X_{1,t}, U_t\big) + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)\\
&= \sum_{t=1}^{n} I\big(U_t, X_{1,t}; Y_{1,t}\big) + I\big(X_{2,t}; Y_{1,t}, Y_{2,t}\big|X_{1,t}, U_t\big) + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)\\
&= \sum_{t=1}^{n} I\big(U_t, X_{1,t}; Y_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{1,t}\big|U_t, X_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t}\big|Y_{1,t}, X_{1,t}, U_t\big)\\
&\qquad + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)\\
&= \sum_{t=1}^{n} I\big(X_{1,t}, X_{2,t}; Y_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t}\big|Y_{1,t}, X_{1,t}, U_t\big) + nD_{21} + n(\epsilon_n^1+\epsilon_n^2)
\end{aligned} \tag{A~33}$$

where the inequality (a) holds because $V_{12}^{L_d}$ is given by a deterministic function of (Y_1^n, Y_2^n) and also conditioning does not reduce the entropy; the inequality (b) is derived by following the same lines as in (A~31) and (A~27). Lastly, we can derive:

$$\begin{aligned}
n(R_1+R_2) &\leq I\big(M_1, M_2; Y_1^n, Y_2^n, V_{21}^{L_d}, V_{12}^{L_d}\big) + n(\epsilon_n^1+\epsilon_n^2)\\
&\overset{a}{=} I(M_1, M_2; Y_1^n, Y_2^n) + n(\epsilon_n^1+\epsilon_n^2)\\
&\leq \sum_{t=1}^{n} I\big(X_{1,t}, X_{2,t}; Y_{1,t}, Y_{2,t}\big) + n(\epsilon_n^1+\epsilon_n^2)
\end{aligned} \tag{A~34}$$

where (a) holds because $V_{21}^{L_d}$ and $V_{12}^{L_d}$ are given by deterministic functions of (Y_1^n, Y_2^n). By collecting (A~24), (A~25), (A~26), (A~29), (A~32), (A~33), (A~34) and applying a standard time-sharing argument, we derive desired constraints of (99) including those given by the auxiliary random variable U. The remaining constraints of (99) can be indeed derived symmetrically. The proof is thus complete. ■

Appendix D
Proof of Theorem 6.6

Consider a length-n code with vanishing average error probability for the Gaussian IC (9) with conferencing decoders. Consider also the auxiliary random variables defined in (A~23). We need to derive the constraints in (113). Define the genie signals $G_{1,t}, G_{2,t}, \tilde{G}_{1,t}$, and $\tilde{G}_{2,t}$ as follows:

$$\begin{cases} G_{1,t} \triangleq a_{21,t}X_{1,t} + Z_{2,t} \\ G_{2,t} \triangleq a_{12,t}X_{2,t} + Z_{1,t} \\ \tilde{G}_{1,t} \triangleq a_{21,t}X_{1,t} + \tilde{Z}_{2,t} \\ \tilde{G}_{2,t} \triangleq a_{12,t}X_{2,t} + \tilde{Z}_{1,t} \end{cases} \qquad t = 1, \dots, n \tag{A~35}$$

where $\{\tilde{Z}_{1,t}\}_{t=1}^{n}$ and $\{\tilde{Z}_{2,t}\}_{t=1}^{n}$ are zero-mean unit-variance Gaussian random processes which are independent of all other random variables. Based on Fano's inequality we have:

$$\begin{aligned}
&n(R_1 + R_2) \\
&\leq I\left(M_1; Y_1^n, V_{21}^{L_d}\right) + I\left(M_2; Y_2^n, V_{12}^{L_d}\right) + n(\epsilon_n^1 + \epsilon_n^2) \\
&\leq I\left(M_1; Y_1^n, G_1^n, V_{21}^{L_d}\right) + I\left(M_2; Y_2^n, G_2^n, V_{12}^{L_d}\right) + n(\epsilon_n^1 + \epsilon_n^2) \\
&= I(M_1; Y_1^n, G_1^n) + I(M_2; Y_2^n, G_2^n) + I\left(M_1; V_{21}^{L_d} \middle| Y_1^n, G_1^n\right) + I\left(M_2; V_{12}^{L_d} \middle| Y_2^n, G_2^n\right) + n(\epsilon_n^1 + \epsilon_n^2) \\
&\leq I(X_1^n; Y_1^n, G_1^n) + I(X_2^n; Y_2^n, G_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&= I(X_1^n; G_1^n) + I(X_2^n; G_2^n) + I(X_1^n; Y_1^n | G_1^n) + I(X_2^n; Y_2^n | G_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&\stackrel{a}{=} H(G_1^n) - H(G_1^n | X_1^n) + H(Y_1^n | G_1^n) - H(Y_1^n | X_1^n) + H(G_2^n) - H(G_2^n | X_2^n) + H(Y_2^n | G_2^n) \\
&\qquad\qquad -H(Y_2^n | X_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&= H(G_1^n) - H(Z_2^n) + H(Y_1^n | G_1^n) - H(G_2^n) + H(G_2^n) - H(Z_1^n) + H(Y_2^n | G_2^n) \\
&\qquad\qquad -H(G_1^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&\stackrel{b}{=} H(Y_1^n | G_1^n) - H(Y_1^n | X_1^n, X_2^n) + H(Y_2^n | G_2^n) - H(Y_2^n | X_1^n, X_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&= H(Y_1^n | G_1^n) - H(Y_1^n | X_1^n, X_2^n, G_1^n) + H(Y_2^n | G_2^n) - H(Y_2^n | X_1^n, X_2^n, G_2^n) \\
&\qquad\qquad\qquad\qquad +n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2) \\
&= I(X_1^n, X_2^n; Y_1^n | G_1^n) + I(X_1^n, X_2^n; Y_2^n | G_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2)
\end{aligned}$$

$$\leq \sum_{t=1}^{n} I\big(X_{1,t}, X_{2,t}; Y_{1,t} \big| G_{1,t}\big) + \sum_{t=1}^{n} I\big(X_{1,t}, X_{2,t}; Y_{2,t} \big| G_{2,t}\big) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + \epsilon_n^2)$$

(A~36)

where equality (a) holds because $G_i^n \to X_i^n \to Y_i^n, i = 1,2$ forms a Markov chain, and equality (b) holds because $H(Y_i^n|X_1^n, X_2^n) = H(Z_i^n), i = 1,2$. We next derive constraints on the linear combination of the rates $R_1 + 2R_2$. We can write:

$$n(R_1 + 2R_2) \leq I\big(M_2; Y_2^n, V_{12}^{L_d}\big) + I\big(M_1; Y_1^n, V_{21}^{L_d}\big) + I\big(M_2; Y_2^n, V_{12}^{L_d}\big) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I(M_2; Y_2^n) + I\big(M_2; V_{12}^{L_d} \big| Y_2^n\big) + I(M_1; Y_1^n)$$
$$+ I\big(M_1; V_{21}^{L_d} \big| Y_1^n\big) + I(M_2; Y_2^n) + I\big(M_2; V_{12}^{L_d} \big| Y_2^n\big) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$\leq I(M_2; Y_2^n, M_1) + I(M_1; Y_1^n, G_1^n) + I(M_2; Y_2^n)$$
$$+ n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I(M_2; Y_2^n|M_1) + I(M_1; Y_1^n|G_1^n) + I(X_1^n; G_1^n) + I(X_2^n; Y_2^n)$$
$$+ n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

(A~37)

Then, for the first two mutual information functions on the right side of the last equality in (A~37) we have:

$$I(M_2; Y_2^n|M_1) + I(M_1; Y_1^n|G_1^n)$$

$$= H(Y_2^n|M_1) - H(Y_1^n|M_1) - H(Y_2^n|M_1, M_2) + H(Y_1^n|G_1^n)$$

$$\overset{a}{=} \sum_{t=1}^{n} H\big(Y_{2,t} \big| M_1, Y_2^{t-1}, Y_{1,t+1}^n\big) - \sum_{t=1}^{n} H\big(Y_{1,t} \big| M_1, Y_2^{t-1}, Y_{1,t+1}^n\big) - H(Y_2^n|M_1, M_2) + H(Y_1^n|G_1^n)$$

$$\overset{b}{\leq} \sum_{t=1}^{n} H\big(Y_{2,t} \big| M_1, Y_2^{t-1}, Y_{1,t+1}^n\big) - \sum_{t=1}^{n} H\big(Y_{1,t} \big| M_1, Y_2^{t-1}, Y_{1,t+1}^n\big)$$
$$- \sum_{t=1}^{n} H\big(Y_{2,t} \big| M_1, M_2, Y_2^{t-1}, Y_{1,t+1}^n\big) + \sum_{t=1}^{n} H\big(Y_{1,t} \big| G_{1,t}\big)$$

$$\overset{c}{=} \sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big)$$

(A~38)

where (a) is derived by [35, Lemma 2]; (b) holds because $Y_2^{t-1}, Y_{1,t+1}^n \to M_1, M_2 \to Y_{2,t}$ forms a Markov chain and conditioning does not reduce the entropy; (c) holds because $X_{i,t}$ is given by a deterministic function of M_i and $G_{1,t} \to U_t, X_{1,t} \to Y_{1,t}$ forms a Markov chain. Now by substituting (A~38) in (A~37), we obtain:

$$
\begin{aligned}
&n(R_1 + 2R_2) \\
&\quad \le \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big) \\
&\qquad\qquad\qquad + I(X_1^n; G_1^n) + I(X_2^n; Y_2^n) + n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad = \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big) \\
&\qquad\qquad\qquad + H(G_1^n) - H(Z_2^n) + H(Y_2^n) - H(G_1^n) + n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad = \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big) \\
&\qquad\qquad\qquad + I(X_1^n, X_2^n; Y_2^n) + n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad \le \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big) \\
&\qquad\qquad\qquad + \textstyle\sum_{t=1}^{n} I\big(X_{1,t}, X_{2,t}; Y_{2,t}\big) + n(2D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)
\end{aligned}
\tag{A~39}
$$

We can also derive:

$$
\begin{aligned}
&n(R_1 + 2R_2) \\
&\quad \le I\big(M_2; Y_2^n, V_{12}^{L_d}\big) + I\big(M_1; Y_1^n, V_{21}^{L_d}\big) + I\big(M_2; Y_2^n, V_{12}^{L_d}\big) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad \le I(M_2; Y_1^n, Y_2^n) + I(M_1; Y_1^n) + I\big(M_1; V_{21}^{L_d} \big| Y_1^n\big) + I(M_2; Y_2^n) + I\big(M_2; V_{12}^{L_d} \big| Y_2^n\big) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad \le I(M_2; Y_1^n, Y_2^n, M_1) + I(M_1; Y_1^n, G_1^n) + I(M_2; Y_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad = I(M_2; Y_1^n | Y_2^n, M_1) + I(M_2; Y_2^n | M_1) + I(M_1; Y_1^n | G_1^n) \\
&\qquad\qquad + I(X_1^n; G_1^n) + I(X_2^n; Y_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad \overset{a}{\le} \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{1,t} \big| Y_{2,t}, X_{1,t}, M_1, Y_2^{t-1}, Y_{1,t+1}^n, Y_{2,t+1}^n\big) \\
&\qquad\qquad + \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big) \\
&\qquad\qquad + I(X_1^n; G_1^n) + I(X_2^n; Y_2^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2) \\
&\quad = \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{1,t} \big| Y_{2,t}, X_{1,t}, U_t, Y_{2,t+1}^n\big) + \textstyle\sum_{t=1}^{n} I\big(X_{2,t}; Y_{2,t} \big| U_t, X_{1,t}\big) + I\big(U_t, X_{1,t}; Y_{1,t} \big| G_{1,t}\big)
\end{aligned}
$$

$$+H(G_1^n) - H(Z_2^n) + H(Y_2^n) - H(G_1^n) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$\overset{b}{\leq} \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t} \middle| Y_{2,t}, X_{1,t}, U_t\right) + \sum_{t=1}^n I\left(X_{2,t}; Y_{2,t} \middle| U_t, X_{1,t}\right) + I\left(U_t, X_{1,t}; Y_{1,t} \middle| G_{1,t}\right)$$

$$+ \sum_{t=1}^n I\left(X_{1,t}, X_{2,t}; Y_{2,t}\right) + n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= \sum_{t=1}^n I\left(X_{2,t}; Y_{1,t}, Y_{2,t} \middle| U_t, X_{1,t}\right) + \sum_{t=1}^n I\left(U_t, X_{1,t}; Y_{1,t} \middle| G_{1,t}\right) + \sum_{t=1}^n I\left(X_{1,t}, X_{2,t}; Y_{2,t}\right)$$

$$+n(D_{12} + D_{21}) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

(A~40)

where (a) is due to (A~38), and (b) holds because conditioning does not reduce the entropy. Lastly, we can write:

$$n(R_1 + 2R_2) \leq I\left(M_1, M_2; Y_1^n, Y_2^n, V_{12}^{L_d}, V_{21}^{L_d}\right) + I\left(M_2; Y_2^n, V_{12}^{L_d}\right) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I(M_1, M_2; Y_1^n, Y_2^n) + I(M_2; Y_2^n) + I\left(M_2; V_{12}^{L_d} \middle| Y_2^n\right) + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$\leq I\left(X_1^n, X_2^n; Y_1^n, Y_2^n, \tilde{G}_1^n\right) + I(X_2^n; Y_2^n) + nD_{12} + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I\left(X_1^n, X_2^n; Y_1^n, Y_2^n \middle| \tilde{G}_1^n\right) + I\left(X_1^n, X_2^n; \tilde{G}_1^n\right) + I(X_2^n; Y_2^n) + nD_{12} + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I\left(X_1^n, X_2^n; Y_1^n, Y_2^n \middle| \tilde{G}_1^n\right) + H\left(\tilde{G}_1^n\right) - H\left(\tilde{Z}_2^n\right)$$

$$+H(Y_2^n) - H(G_1^n) + nD_{12} + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$= I\left(X_1^n, X_2^n; Y_1^n, Y_2^n \middle| \tilde{G}_1^n\right) + I(X_1^n, X_2^n; Y_2^n) + nD_{12} + n(\epsilon_n^1 + 2\epsilon_n^2)$$

$$\leq \sum_{t=1}^n I\left(X_{1,t}, X_{2,t}; Y_{1,t}, Y_{2,t} \middle| \tilde{G}_{1,t}\right) + \sum_{t=1}^n I\left(X_{1,t}, X_{2,t}; Y_{2,t}\right) + nD_{12} + n(\epsilon_n^1 + 2\epsilon_n^2)$$

(A~41)

Finally, by applying a standard time-sharing argument to (A~36), (A~39), (A~40), and (A~41), we derive $1^{st}, 3^{rd}, 5^{th}$, and 7^{th} constraint of (113), respectively. The remaining constraints of (113) could be symmetrically derived (similar to (A~39), (A~40), and (A~41)). The proof is thus complete. ■