Approximate Capacity of Gaussian Interference-Relay Networks with Weak Cross Links

S. Mohajer S N. Diggavi C. Fragouli D N C. Tse

Abstract

In this paper we study a Gaussian relay-interference network, in which relay (helper) nodes are to facilitate competing information flows over a wireless network. We focus on a two-stage relay-interference network where there are weak cross-links, causing the networks to behave like a chain of Z Gaussian channels. For these Gaussian ZZ and ZS networks, we establish an approximate characterization of the rate region. The outer bounds to the capacity region are established using genie-aided techniques that yield bounds sharper than the traditional cut-set outer bounds. For the inner bound of the ZZ network, we propose a new interference management scheme, termed interference neutralization, which is implemented using structured lattice codes. This technique allows for over-the-air interference removal, without the transmitters having complete access the interfering signals. For both the ZZ and ZS networks, we establish a new network decomposition technique that (approximately) achieves the capacity region. We use insights gained from an exact characterization of the corresponding linear deterministic version of the problems, in order to establish the approximate characterization for Gaussian networks.

I. Introduction

The multi-commodity flow problem, where multiple independent unicast sessions need to share network resources, can be solved efficiently over graphs using linear programming techniques [1]. This is not the case for wireless networks, where the broadcast and superposition nature of the wireless medium introduces complex signal interactions between the competing flows. The simplest example is the one-hop interference channel [2], where two transmitters with independent messages are attempting to communicate with their respective receivers over the wireless transmission medium. Even for this simple one-hop network, the information-theoretic characterization has been open for several decades. To

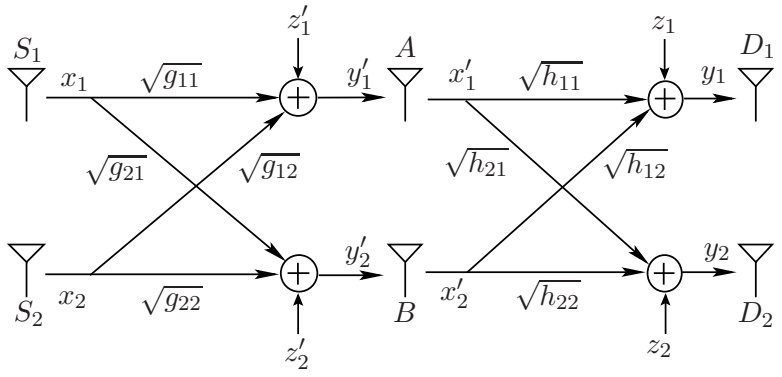


Fig. 1. Two-stage relay-interference network.

study more general networks, there is a clear need to understand and develop sophisticated interference management techniques.

The capacity of the wireless Gaussian interference channel has been (approximately) characterized, within one bit (see [3] and the references therein). Building on this progress, a natural next step is to study the approximate capacity region of small-scale *interference-relay networks*, where there are potentially multiple hops from the sources to destinations through cooperating relays. Studying even simple two-hop topologies could help develop techniques and build insight that would enable a (perhaps approximate) characterization of capacity for more general networks. We are interested in our work in a *universal* type of approximation, in that it should characterize the capacity to within a constant number of bits, independently of the signal-to-noise ratio and the channel parameter values.

The focus of this paper is to study the two-stage relay-interference network illustrated in Figure 1. In particular, we give an approximate characterization of the capacity region for special cases of these networks when some of the cross-links are weak. These are illustrated in Figure 3(a) and Figure 4(a), which we refer to as the ZS and ZZ Gaussian models. We first study a *deterministic* version of these problems by using the linear deterministic model introduced in [4]. An exact capacity region characterization in the deterministic case is then translated into a universally approximate characterization for the (noisy) Gaussian network. In particular, for ZS and ZZ networks we have a capacity region characterization within 2 bits (or less), independent of the operating signal-to-noise ratio and the channel parameters.

In studying these special networks, we discover that many sophisticated techniques are required to (approximately) characterize the network capacity region. The main new ingredients that enable this characterization are as follows: (i) a new interference management technique we term *interference neutralization*, in which interference is canceled over the air, without the relays necessarily decoding the

transmitted messages[1]; (ii) a *structured lattice code* that enables interference neutralization over Gaussian networks; (iii) A *network decomposition* technique which enables appropriate rate-splitting of the message and power allocation for the different message components; (iv) genie-aided outer bounding techniques that enable bounds that are tighter than the information-theoretic cut-set outer bounds.

A way to interpret the achievability results for the ZS networks is that the relays perform a partial-decoding of strategically split messages from the sources, and then cooperate to deliver the required messages to the destination, again through strategically splitting the messages. The power allocated to each of the sub-messages is determined using the insight derived from the deterministic model, that messages that are not intended be decoded arrive at the noise-level. The achievability for the ZZ network is slightly more sophisticated in that one of the relays is required to only decode a function of the sub-messages. The function is chosen such that its signal in combination with the transmission of the other relay causes the unwanted interference to be cancelled (neutralized) at the destination. This interference neutralization is enabled in the Gaussian channel using the group property of a structured lattice code.

Work in the literature over the past decade has examined scaling laws for multiple independent flows over wireless networks, see for example [6]–[8]. The goal there is to characterize the order of the wireless network capacity as the network size grows. In contrast, in our work, instead of seeking order arguments and scaling laws, we try to characterize the capacity (perhaps within a universal constant of a few bits) for specific topologies. The interference channel is a special case of such networks, where there is only one-hop communication between the sources and destinations. There has been a surge of recent work on this topic including cooperating destinations [9] and use of feedback in inducing cooperation at the transmitters [10]. The deterministic approach developed in [4] has been successfully applied to the interference channel in [11]. The fundamental role of interference alignment in K-user interference channel (still a one-hop network) has been demonstrated in [12], [13].

The paper is organized as follows. Section II introduces our notation and the basic network models we study. Section III illustrates the transmission techniques used in this paper through simple deterministic examples. The main results are given in Section IV, along with a proof outline for the Gaussian networks in Section V. The achievability and converse for the deterministic ZS network is given in Section VI, and many of these ideas are translated into the precise proof for the corresponding Gaussian ZS network in

[1]A noise nulling technique is proposed in [5] to mitigate correlated noise in an amplify-forward relaying strategy for a single unicast "diamond" parallel relay network. However, the difference in our technique is that we use the structure of the codebooks (without necessarily decoding information) to neutralize interference, and not noise statistics. Moreover the multiple-unicast nature of the problem necessitates strategic partitioning and rate-splitting of different components of the messages.

Appendix A. Section VII follows a similar program for the ZZ network, by first identifying the capacity region for the deterministic version. This allows illustration of ideas such as interference neutralization, as well as genie-aided outer bounding techniques. The precise translation of these results into Gaussian ZZ networks is given in Appendix B. Section VIII concludes the paper with a short discussion.

II. Problem Statement

A well accepted model for wireless communication is a linear Gaussian model. In this, the received signal $y_i(t)$ at time t, is related to the transmitted signals $\{x_j[t]\}$ as

$$y_i[t] = \sum_j h_{ij} x_j[t] + z_i[t], \tag{1}$$

where $z_i(t)$ is i.i.d. (unit-variance) Gaussian noise, and h_{ij} represents the fading channel from transmitter i to receiver j.

II.1 The Deterministic Model

In [4], a deterministic model was proposed, to capture the essence of wireless interaction described in (1). The advantage of the deterministic model is its simplicity, which allows exact characterizations; its purpose is to build insights for the noisy wireless network in (1). The deterministic model of [4] simplifies the wireless interaction model by eliminating the noise and discretizing the channel gains through a binary expansion of q bits. Therefore, the received signal Y_i, which is a binary vector of size q, is modeled as

$$Y_i[t] = \sum_j N_{ij} X_j[t], \tag{2}$$

where N_{ij} is a $q \times q$ binary matrix representing the (discretized) channel transformation between nodes j and i and X_j is a $q \times 1$ vector that contains the (discretized) transmitted signal. We will drop the time index t when it does not play a role for simplicity. All operations in (2) are done over the binary field, $\mathbb{F}_2$. We use the terminology *deterministic wireless network* when the signal interaction model is governed by (2). The model in (2) is an approximate representation of a Gaussian fading channel, which attempts to capture the attenuation effect of the signal caused by the channel gain. This can be interpreted as the number of significant bits of a binary representation of the input, x_j, that is above the noise level. More

precisely, typically the model in (2) assigns $N_{ij} = \mathbf{J}^{q-n_{ij}}$, where $\mathbf{J}$ is a shift matrix, i.e.,

$$\mathbf{J} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}_{q\times q} . \tag{3}$$

For real channel gain h_{ij} in the Gaussian model (1), we calculate n_{ij} as $n_{ij} = \lceil \frac{1}{2}\log|h_{ij}|^2 \rceil$. The parameter q is chosen such that $q \geq \max_{i,j} \lceil \frac{1}{2}\log|h_{ij}|^2 \rceil$.

An example of a deterministic network is illustrated in Figure 2. Each node contains several channel inputs and outputs, which are called *sub-node* or *level* through out this paper. Source S_1 can only send one bit to node A and no bit to node B; source S_2 can send its two MSB to both A and B, and its LSB to node B. The transmitted bits from nodes S_1 and S_2 interfere on the LSB that node A receives.

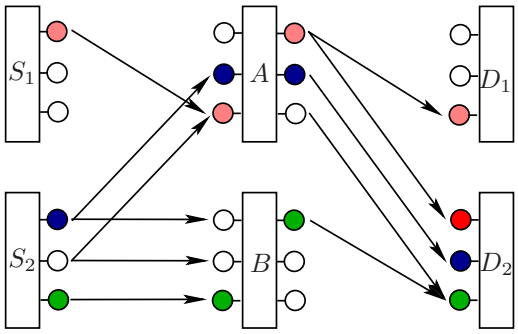


Fig. 2. A deterministic network.

In a deterministic network, given a cut that separates nodes $\mathcal{U}$ from node $\mathcal{V}$, the cut-value equals the rank of the transfer matrix between the nodes in $\mathcal{U}$ and $\mathcal{V}$. For example, in Figure 2, the cut that separates nodes $\mathcal{U} = \{S_1, S_2, A, B\}$ and $\mathcal{V} = \{D_1, D_2\}$ equals

$$\text{rank}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 3. \tag{4}$$

The rows of this transfer matrix correspond to the four transmitted inputs by nodes A and B, while the columns to the four receives outputs at nodes D_1 and D_2.

II.2 Interference-relay network model

Our goal in this paper is to derive approximate capacity characterizations for a class of 2-user relay-interference networks shown in Figure 1, which we call the XX network. We start by describing our

notation for Gaussian channels.

Two transmitters, S_1 and S_2, encode their messages W_1 and W_2 of rates R_1 and R_2, respectively, and broadcast the obtained signals to the relay nodes, A and B. Denote the transmitted signals by x_1 and x_2, and the received signals at the relays by y_1' and y_2'. Then

$$\begin{aligned} y_1'[t] &= \sqrt{g_{11}}x_1[t] + \sqrt{g_{12}}x_2[t] + z_1'[t], \\ y_2'[t] &= \sqrt{g_{21}}x_1[t] + \sqrt{g_{22}}x_2[t] + z_2'[t], \end{aligned} \tag{5}$$

where z_1', z_2' are unit-variance Gaussian noises, independent of each other and of x_1, x_2.

The relay nodes perform any (causal) processing on their received signal sequences $\{y_1'[t]\}$ and $\{y_2'[t]\}$ respectively, to obtain their transmitting signal sequences, $\{x_1'(t)\}$ and $\{x_2'(t)\}$. The received signals at the destination nodes can be written as

$$\begin{aligned} y_1[t] &= \sqrt{h_{11}}x_1'[t] + \sqrt{h_{12}}x_2'[t] + z_1[t] \\ y_2[t] &= \sqrt{h_{21}}x_1'[t] + \sqrt{h_{22}}x_2'[t] + z_2[t], \end{aligned} \tag{6}$$

where the z_1', z_2', z_1, and z_2 are independent zero-mean unit-variance noises, which are also independent of x_1 and x_2. There is a power constraint for each transmitted signal, that is, $\mathbb{E}[x_1^2] \leq 1$, $\mathbb{E}[x_2^2] \leq 1$, $\mathbb{E}[x_1'^2] \leq 1$ and $\mathbb{E}[x_2'^2] \leq 1$.

Each destination node D_i, $i = 1, 2$, is interested in decoding its message W_i, using its received signals $\{y_i[t]\}$. We define a rate pair (R_1, R_2) to be admissible if there exist a transmission scheme under which D_1 and D_2 can decode W_1 and W_2, respectively, with arbitrary small (average) error probability in the standard manner [14]. This would allow two end-to-end reliable unicast sessions at rates (R_1, R_2) for the source/destination pairs (S_1, D_1) and (S_2, D_2).

A useful tool to examine the network problem defined above is to study its deterministic version, based on the model developed in (2). Using the deterministic approach, we can rewrite (5)-(6) as

$$\begin{aligned} Y_1'[t] &= M_{11}X_1[t] + M_{12}X_2[t] \\ Y_2'[t] &= M_{21}X_1[t] + M_{22}X_2[t], \end{aligned} \tag{7}$$

and

$$\begin{aligned} Y_1[t] &= N_{11}X_1'[t] + N_{12}X_2'[t] \\ Y_2[t] &= N_{21}X_1'[t] + N_{22}X_2'[t], \end{aligned} \tag{8}$$

where the matrices $\{M_{ij}\}$ and $\{N_{ij}\}$ approximately model the channels in (5)-(6), *i.e.*, $M_{ij} = \mathbf{J}^{q-m_{ij}}, N_{ij} = \mathbf{J}^{q-n_{ij}}$. The matrix $\mathbf{J}$ is defined as in (3), while $m_{ij} = \lceil \frac{1}{2}\log|g_{ij}|^2 \rceil$ and $n_{ij} = \lceil \frac{1}{2}\log|h_{ij}|^2 \rceil$.

It is worth mentioning that though this network looks like cascaded interference channels, there is an important difference. Unlike the interference channel, the messages sent by the relays at the second layer

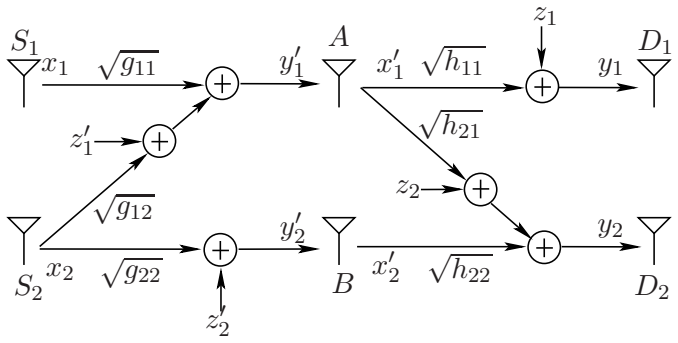
of transmission need not independent, *i.e.,* we can try to induce cooperation at the relays to transmit information to the final destinations. This distinction makes this network more interesting than a simple cascade of interference channels.

In this paper, we focus on two specific realizations of the network, namely, the ZS and the ZZ networks, which further simplify the connectivity models of (5)-(6). We describe these two networks in the following, and give an approximate characterization of their admissible rate region in Section IV.

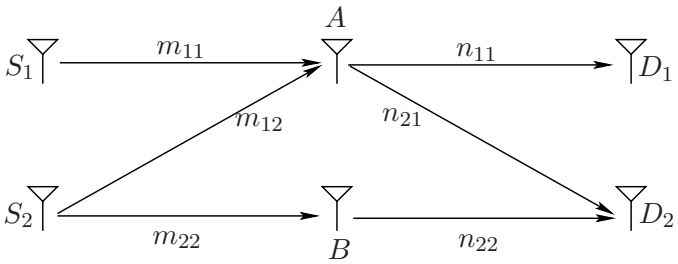
Notation alert: Throughout this paper, we use the lowercase letters x and y for the signals transmitted by the sources and received signals at the destinations in the Gaussian networks. The received and transmitting signals by the relays are denoted by x' and y'. Similarly, uppercase letters will be used for the deterministic networks.

II.3 The ZS Network

The ZS network is a special case of the interference-relay network defined in (5)-(6). In the ZS network one cross link in each layer has a negligible gain, and therefore does not cause interference, as illustrated in Figure 3(a). In particular, we assume $g_{21} = h_{12} = 0$ in the Gaussian network, and $m_{21} = n_{12} = 0$ in the deterministic network. The resulting Gaussian ZS network is shown in Figure 3(a), and the deterministic model for this network is given in Figure 3(b).



(a) The Gaussian ZS network

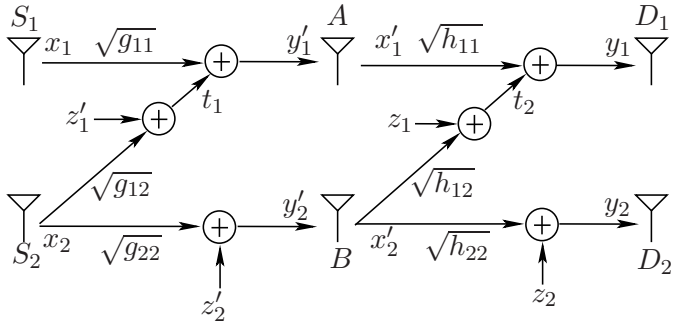


(b) The deterministic ZS network

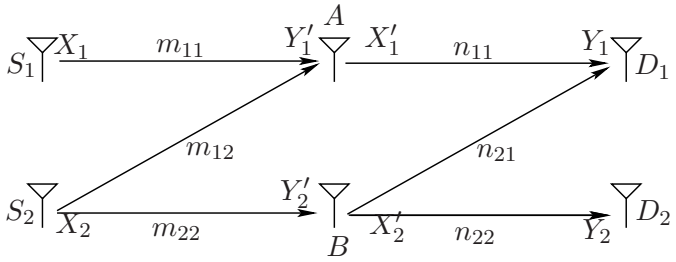
Fig. 3. The ZS network.

II.4 The ZZ Network

The ZZ network is another special configuration interference-relay network, wherein one cross link in each layer has zero gain. However, the difference is that, here the missing links are in parallel. In particular, we assume $g_{21} = h_{21} = 0$ and $m_{21} = n_{21} = 0$ in the Gaussian and deterministic networks, respectively. The Gaussian and corresponding deterministic ZZ networks are shown in Figure 4.



(a) The Gaussian ZZ network



(b) The deterministic ZZ network

Fig. 4. The ZZ network.

III. Examples illustrating transmission techniques

In this section, we illustrate through examples some of the main interference management techniques we will use to (approximately) achieve the capacity of our relay-interference networks. For simplicity in demonstrating the ideas, we focus on deterministic networks throughout the examples. However, similar techniques will be used later for Gaussian channels as well.

We also present a simple Gaussian example at the end of this section, to illustrate the message splitting idea used in many places through out this work.

Example 1 (Network Decomposition for the ZS Network): A deterministic ZS network can be always decomposed into two subnode-disjoint networks, where the first partition consists of a set of sub-nodes of S_1, A and D_1, and looks like a line network. The second partition is, however, a diamond network, with a broadcast channel from S_2 to A and B in the first layer, and a multiple access channel from A

and B to D_2 in the second layer. This diamond network can be used to send information from S_2 to D_2. Since these two networks are sub-node disjoint, there would be no interfering signal, and each of them can be analyzed separately. This is more illustrated in Figure 5.

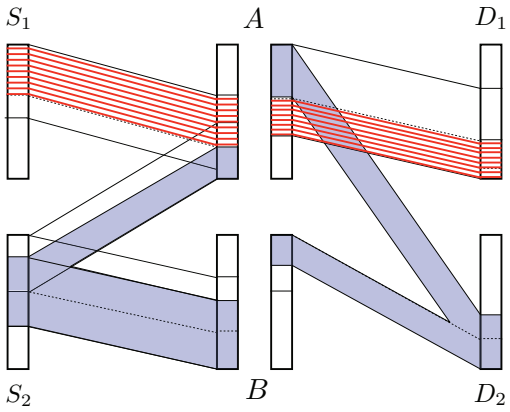


Fig. 5. Network partitioning for a deterministic ZS network.

In a Gaussian ZS network the network decomposition can be done using message splitting, superposition coding and proper power allocation. We will use this technique to achieve an approximate capacity for the Gaussian ZS network.

Example 2 (Interference Neutralization): This technique can be used in networks which contain more than one disjoint path from S_i to D_j for $i \neq j$, where D_j is not interested in decoding the message sent by the source node S_i, and therefore it receives the interference through more than one link. The proposed technique is to tune these interfering signals such that they neutralize each other at the destination node. In words, the interfering signal should be received at the same power level and with different sign such that the effective interference, obtained by adding them, occupies a smaller number of degrees of freedom. To best of our knowledge, this technique is new and was been introduced in [15].

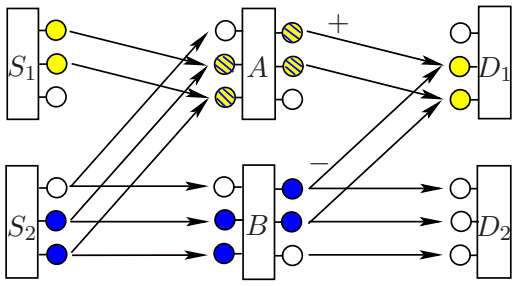


Fig. 6. Interference Neutralization; $(R_1, R_2) = (2, 2)$ is achievable.

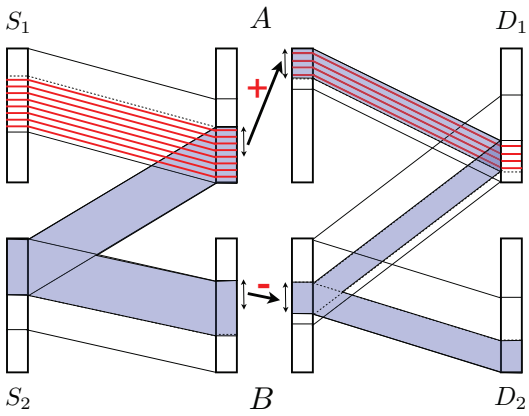


Fig. 7. Interference neutralization.

Figure 6 shows a network in which interference neutralization is essential to achieve the desired rate pair $(R_1, R_2) = (2, 3)$. Here D_1 has only two degrees of freedom, and receives information bits from both A and B over these sub-nodes. However, notice that there are two disjoint paths (S_2, A, D_1) and (S_2, B, D_1), which connect S_2 to D_1. As it is shown in Figure 6, using a proper mapping (permutation) at the relay nodes, one can make the interference neutralized at the destination node D_1, and provide two non-interfered links from S_1 to D_2. Note that this permutation does not effect the admissible rate of the other unicast from S_2 to D_2, the cost we pay, is to permute the received bits at D_2. A more general illustration of this phenomenon is given in Figure 7.

Example 3 (Use of Lattice Codes to Implement Interference Neutralization over Gaussian ZZ *Network):* The idea of interference neutralization illustrated in Example 2 can be also used in Gaussian networks. In this case a group structured code, such as lattice code, is required to play the role of composition and decomposition of the signal and interference in two layers of the network. Consider the Gaussian ZZ network in Figure 4(a). We can use message splitting and interference neutralization to improve the achievable rate pairs of this network.

Let the second source split its message into two parts as $W_2 = (W_2^{(N)}, W_2^{(P)})$, namely, the functional (neutralization) and private parts, of rates $R_{2,N} = R_1$ and $R_{2,P} = R_2 - R_1$. Both transmitters use a common lattice code to encode W_1 and $W_2^{(N)}$, and map them into $\mathbf{x}_1^{(N)}$ and $\mathbf{x}_2^{(N)}$, respectively. The other message $W_2^{(P)}$ can be encoded to $\mathbf{x}_2^{(P)}$ using a random Gaussian code. We assume that both the lattice code and the random Gaussian code have average power equal to 1. Then, the transmitting signals would be a linear combination of the codewords with a proper power allocation, *i.e.*,

$$\mathbf{x}_1 = \sqrt{\alpha_N}\mathbf{x}_1^{(N)}, \qquad \mathbf{x}_2 = \sqrt{\beta_N}\mathbf{x}_2^{(N)} + \sqrt{\beta_P}\mathbf{x}_2^{(P)}, \tag{9}$$

where the power allocation coefficients satisfy $\alpha_N \leq 1$ and $\beta_N + \beta_P \leq 1$. The transmitters choose the power allocated to $\mathbf{x}_1^{(N)}$ to $\mathbf{x}_2^{(N)}$ in a way that they get received at A with the same power. In this way, their summation would be again a lattice code and can be decoded at A by treating $\mathbf{x}_1^{(P)}$ as noise. A similar strategy will be used for signaling at the relay for transmission in the second layer of the network. The only difference is that instead of sending $\mathbf{x}_2^{(N)}$, the relay node B sends $-\mathbf{x}_2^{(N)}$. Then, the lattice point observed at D_1 would be exactly $\mathbf{x}_1^{(N)}$ and it can find W_1. The other decoder can simply first reverse $-\mathbf{x}_2^{(N)}$ to $\mathbf{x}_2^{(N)}$, and then decode it. This idea is illustrated in Figure 8.

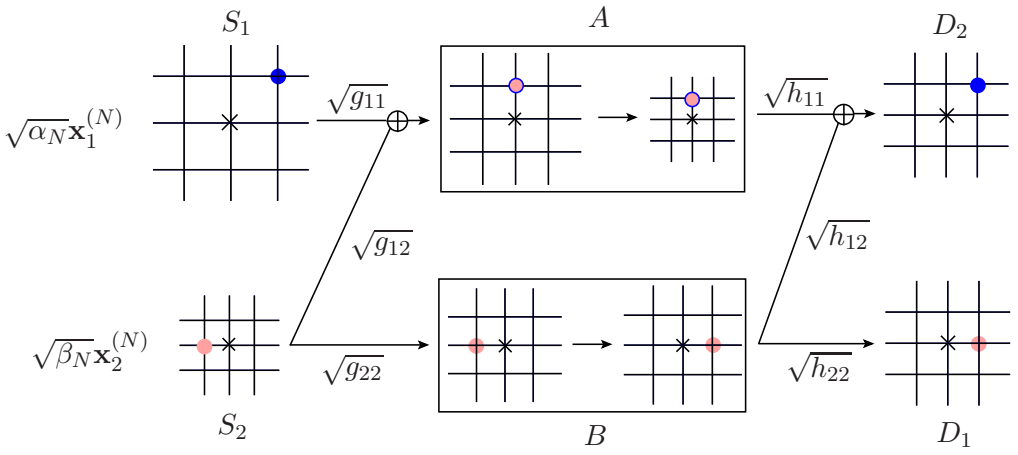


Fig. 8. Using lattice codes for interference neutralization over a Gaussian ZZ network. The origin is specified by a cross "$\times$". Power allocated to the messages at the transmitters are chosen such that the two lattice points corresponding to $\mathbf{x}_1^{(N)}$ and $\mathbf{x}_2^{(N)}$ get received at B at the same power level, and their summation becomes a point on the scaled lattice. The same strategy is used by the relays. The relay node B also reverses its transmitting lattice point in order to neutralize the interference caused in the first layer of the network.

Example 4: Consider the Gaussian Z network shown in Figure 9, with channel gains $g_{11} \geq 1$, $g_{12} \geq 1$, and $g_{22} \geq 1$.

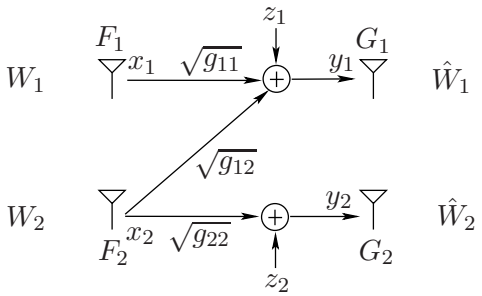


Fig. 9. A Gaussian Z network.

The source nodes F_i wishes to encode and send message W_i to the destination node G_i, for $i = 1, 2$.

Denoting the rate of message W_i by R_i, an approximate capacity characterization for this network is given by network is given by

$$\begin{aligned}\mathcal{R}^{\mathsf{Z}} = \Big\{(R_1,R_2) : & R_1 \le \frac{1}{2}\log\left(1+g_{11}\right) \\ & R_2 \le \frac{1}{2}\log\left(1+g_{22}\right) \\ & R_1+R_2 \le \frac{1}{2}\log\left(1+g_{11}+g_{12}\right)+\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right)\Big\}.\end{aligned} \tag{10}$$

It is easy to show that any achievable rate pair belongs to $\mathcal{R}^{\mathsf{Z}}$, and hence $\mathcal{R}^{\mathsf{Z}}$ establishes an outer bound for the capacity region. Moreover, one can show that the rate pair $(R_1-\frac{1}{2}, R_2-\frac{1}{2})$ is achievable provided that $(R_1,R_2)\in\mathcal{R}^{\mathsf{Z}}$. The encoding strategy to achieve such rate pair involves message splitting and proper power allocation. We will discuss this in more details in Appendix C.

IV. Main Results

In this section we present the main results of this paper, which is the approximate capacity characterization of the Gaussian ZS and ZZ interference-relay networks. In order to obtain such an approximate characterization, we have a complete characterization of the deterministic versions of the ZS and ZZ networks. The coding strategies for the Gaussian problems are outlined in Section V. The detailed analysis of these strategies and the corresponding outer bounds which lead to Theorems 2 and 4 are given in Appendices A and B, respectively. Most of the insights are obtained by analyzing the deterministic versions of these problems, and the exact characterizations are summarized in Theorems 1 and 3 respectively. We prove these results in Sections VI and VII, respectively. The achievability and outer bound results for the Gaussian cases are directly inspired by these results.

IV.1 The ZS Network

The ZS network illustrated in Figure 3(b) and the corresponding Gaussian ZS network is given in Figure 3(a). Theorems 1 and 2 give the exact and approximate (within 2 bits) characterizations of their capacity regions.

Theorem 1 (The capacity region of deterministic ZS network): The capacity region of the determinis-

tic ZS network is specified by $\mathcal{R}^{\mathsf{DZS}}$, where $\mathcal{R}^{\mathsf{DZS}}$ is the set of all rate pairs (R_1, R_2) that satisfy

$$R_1 \leq m_{11}, \tag{DZS-1}$$

$$R_2 \leq \max(m_{12}, m_{22}), \tag{DZS-2}$$

$$R_1 + R_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+, \tag{DZS-3}$$

$$R_2 \leq m_{12} + n_{22}, \tag{DZS-4}$$

$$R_1 + R_2 \leq m_{22} + \max(n_{11}, n_{21}), \tag{DZS-5}$$

$$R_1 + R_2 \leq \max(m_{11}, m_{12}) + n_{22}, \tag{DZS-6}$$

$$R_1 \leq n_{11}, \tag{DZS-7}$$

$$R_2 \leq \max(n_{21}, n_{22}), \tag{DZS-8}$$

$$R_2 \leq m_{22} + n_{21}, \tag{DZS-9}$$

$$R_1 + R_2 \leq \max(n_{21}, n_{22}) + (n_{11} - n_{21})^+. \tag{DZS-10}$$

Theorem 2 (An Approximate capacity region of Gaussian ZS network): Let $\mathcal{R}^{\mathsf{GZS}}$ be the set of all rate pairs (R_1, R_2) which satisfy (GZS-1)–(GZS-10) given below. Then $\mathcal{R}^{\mathsf{GZS}}$ is an outer bound for the capacity region of the Gaussian ZS network. Moreover, for any $(R_1, R_2) \in \mathcal{R}^{\mathsf{GZS}}$, there exists a transmission scheme with rates $(R_1', R_2') = (R_1 - \delta_1, R_2 - \delta_2)$, where $\delta_1 = 1$ and $\delta_2 = 1.5$ are universal

constants, independent of the channel gain, and required rates.

$$R_1 \leq \frac{1}{2}\log(1+g_{11}) \tag{GZS-1}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{12}+g_{22}) \tag{GZS-2}$$

$$R_1+R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12})+\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) \tag{GZS-3}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{12})+\frac{1}{2}\log(1+h_{22}) \tag{GZS-4}$$

$$R_1+R_2 \leq \frac{1}{2}\log(1+g_{22})+\frac{1}{2}\log(1+h_{11}+h_{21}) \tag{GZS-5}$$

$$R_1+R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12})+\frac{1}{2}\log(1+h_{22}) \tag{GZS-6}$$

$$R_1 \leq \frac{1}{2}\log(1+h_{11}) \tag{GZS-7}$$

$$R_2 \leq \frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}}) \tag{GZS-8}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{22})+\frac{1}{2}\log(1+h_{21}) \tag{GZS-9}$$

$$R_1+R_2 \leq \frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}})+\frac{1}{2}\log\left(1+\frac{h_{11}}{h_{21}}\right). \tag{GZS-10}$$

The outer bound for the results above are fairly standard arguments based on reducing a multi-letter mutual information into single-letter forms by appropriately using decodability requirements at the different destinations. The details of these are given in Section VI.1 and Appendix B.1, respectively.

The coding strategy achieving these regions is based on two ideas. One is that of a *network decomposition* illustrated in Section III, Example 1 for the deterministic network. The insight from the network decomposition leads to the idea of strategic *rate-splitting* and power allocation in the Gaussian channel. For the Gaussian coding scheme, we need to strategically partition the messages and allocate powers in order for the relays to partially decode appropriate messages and setup cooperation. The details of this strategy are outlined in Section V.

IV.2 The ZZ *Network*

The ZZ network illustrated in Figure 4(b) and the corresponding Gaussian ZZ network is given in Figure 4(a). Although superficially the ZS and ZZ networks may look similar, the subtle difference in the network connectivity, makes the two problems completely different, both in terms of capacity characterization, as well as transmission schemes. It will be shown that a new interference management scheme, which we term as interference neutralization, is needed to (approximately) achieve the capacity

of this network. The most intuitive description for interference neutralization is to cancel interference over air without processing at the destinations. This scheme can be used whenever there are more than one path for interference to get received at a destination. We will explain it in more detail in Sections V and VII.

Theorems 3 and 4 give the exact and approximate (within 2 bits) characterizations for the capacity region of the deterministic and the Gaussian ZZ networks, respectively. Another new ingredien used here is needed a genie-aided outer bound that gives the (noisy) cross link of the first (or correspondingly second) layer to the destination (or correspondingly to the relay). This genie-aided bound allows us to develop outer bounds that are apparantly tighter than the information-theoretic cut-set bounds by utilizing the decoding structure needed.

Theorem 3 (The capacity region of deterministic ZZ *network):* The capacity region of the deterministic ZZ network is given by $\mathcal{R}^{\mathsf{DZZ}}$, where $\mathcal{R}^{\mathsf{DZZ}}$ is the set of all rate pairs (R_1, R_2) which satisfy

$$R_1 \leq m_{11}, \tag{DZZ-1}$$

$$R_2 \leq m_{22}, \tag{DZZ-2}$$

$$R_1 \leq n_{11}, \tag{DZZ-3}$$

$$R_2 \leq n_{22}, \tag{DZZ-4}$$

$$R_1 + r_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ + n_{12}, \tag{DZZ-5}$$

$$R_1 + R_2 \leq \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ + m_{12}. \tag{DZZ-6}$$

Theorem 4 (An approximate capacity region of Gaussian ZZ *network):* Let $\mathcal{R}^{\mathsf{GZZ}}$ be the set of all rate pairs (R_1, R_2) which satisfy (GZZ1)–(GZZ6) given below.

$$R_1 \leq \frac{1}{2}\log(1 + g_{11}) \tag{GZZ1}$$

$$R_2 \leq \frac{1}{2}\log(1 + g_{22}) \tag{GZZ2}$$

$$R_1 \leq \frac{1}{2}\log(1 + h_{11}) \tag{GZZ3}$$

$$R_2 \leq \frac{1}{2}\log(1 + h_{22}) \tag{GZZ4}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) + \frac{1}{2}\log(1 + h_{12}), \tag{GZZ5}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + h_{11} + h_{12}) + \frac{1}{2}\log\left(1 + \frac{h_{22}}{h_{12}}\right) + \frac{1}{2}\log(1 + g_{12}) \tag{GZZ6}$$

Then, any admissible rate pair (R_1, R_2) for the Gaussian ZZ networks belongs to $\mathcal{R}^{\mathsf{GZZ}}$. Moreover, for any rate pair $(R_1, R_2) \in \mathcal{R}^{\mathsf{GZZ}}$, there exists an encoding scheme with rates $(R_1', R_2') = (R_1 - \frac{7}{4}, R_2 - \frac{7}{4})$.

V. Gaussian coding strategies

This section is devoted to providing the basic ideas of the coding schemes used in the Gaussian ZS and ZZ networks. We also develop an outline of how to analyze these coding strategies.

V.1 The Gaussian ZS *network: Achievability*

The coding strategy for the Gaussian ZS network is essentially a partial-decode-and-forward strategy, along with a strategic rate-splitting of the messages. Let the messages to be sent from S_1, S_2 be denoted by W_1, W_2 respectively (see Figure 3(a)). We will break the ZS network into two cascaded interference channels, where we require particular messages to be decoded at the relays and forwarded to the destinations. The first stage is a Z interference channel, where the message W_2 is split into three parts: $\left(U_2^{(1)}, U_2^{(2)}, U_2^{(3)}\right)$. The intention of this strategic split is to allow the the node G_1 (which is relay A in the original ZS network) to decode $\left(U_1^{(1)}, U_2^{(1)}, U_2^{(2)}\right)$ and node G_2 (which is relay B in the original ZS network), to decode $\left(U_2^{(1)}, U_2^{(3)}\right)$. This is illustrated in Figure 10. Here, $U_2^{(1)}$ plays the role of a common message which can be decoded at both receivers, whereas $U_2^{(2)}$ and $U_2^{(3)}$ are the private messages for G_1 and G_2 respectively.

The next stage of the ZS network is a S interference channel depicted in Figure 11. Here we take the messages delivered and decoded by the Z interference channel of the first stage and further process them to ensure delivery of the desired messages to the destination. In particular, we further split the decoded messages from the first stage into several parts and require delivery of messages as shown in Figure 11. This splitting and delivery of appropriate pieces, finally ensures that W_1 and W_2 are decodable at the destinations. This is the encoding strategy in the ZS network. In the following lemmas, we give the rates at which messages at each stage can be delivered. Putting together Lemmas 1 and 2, we get the desired result given in Theorem 2. The proofs of these lemmas follow fairly standard arguments, and are given in Appendix C.

A formal statement of the argument above is given below.

***Lemma** 1:* Consider the Gaussian Z interference network with channel gains (g_{11}, g_{12}, g_{22}), and decoding requirements as shown in Figure 10. Denoting the rate of the sub-message $U_i^{(j)}$ by $\Upsilon_{i,j}$, any rate

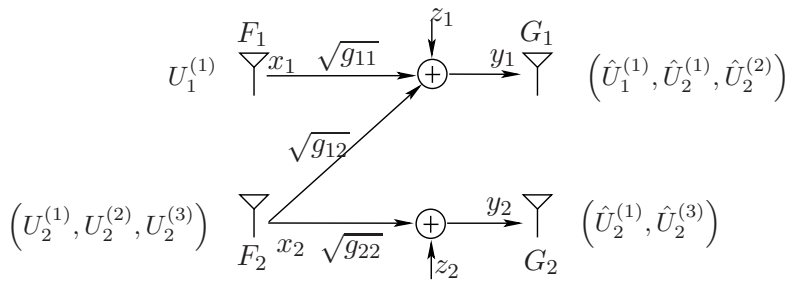


Fig. 10. The Z interference channel with particular message requirements, captures the proposed coding scheme for the first layer of the Gaussian ZS network.

tuple $(\Upsilon_{1,1}, \Upsilon_{2,1}, \Upsilon_{2,2}, \Upsilon_{2,3})$ which satisfies

$$\Upsilon_{1,1} \leq \left(\frac{1}{2}\log\left(1+g_{11}\right) - \frac{1}{2}\right)^+, \tag{7}$$

$$\Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+\frac{g_{12}}{g_{22}}\right) - \frac{1}{2}\right)^+, \tag{8}$$

$$\Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+g_{12}\right) - \frac{1}{2}\right)^+, \tag{9}$$

$$\Upsilon_{1,1} + \Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+g_{11}+g_{12}\right) - \frac{1}{2}\right)^+, \tag{10}$$

$$\Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\right)^+, \tag{11}$$

$$\Upsilon_{2,1} + \Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1+g_{22}\right) - \frac{1}{2}\right)^+, \tag{12}$$

is achievable.

The next lemma gives an achievable rate region for the second layer of the ZS network, which is a S interference network depicted in Figure 11.

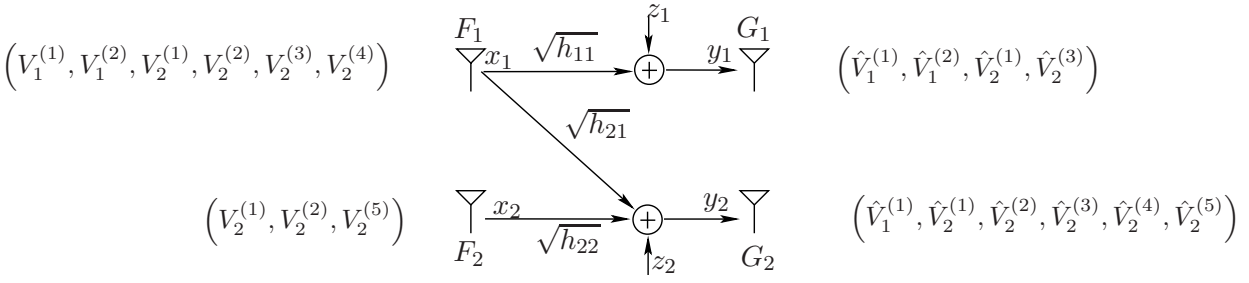


Fig. 11. The S interference channel with particular message requirements, depicting the proposed coding strategy for the second layer of the Gaussian ZS network.

***Lemma** 2:* Consider the Gaussian S interference network with channel gains (h_{11}, h_{21}, h_{22}), and decoding requirements as shown in Figure 11, where $\Theta_{i,j}$ denotes the rate of message $V_i^{(j)}$. Any rate tuple $(\Theta_{1,1}, \Theta_{1,2}, \Theta_{2,1}, \Theta_{2,2}, \Theta_{2,3}, \Theta_{2,4}, \Theta_{2,5})$ which satisfies

$$\Theta_{1,1} + \Theta_{1,2} + \Theta_{2,1} + \Theta_{2,3} \leq \left(\frac{1}{2}\log\left(1 + h_{11}\right) - \frac{1}{2}\right)^+, \tag{13}$$

$$\Theta_{1,2} \leq \left(\frac{1}{2}\log\left(1 + \frac{h_{11}}{h_{12}}\right) - \frac{1}{2}\right)^+, \tag{14}$$

$$\Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1 + \frac{h_{21}}{h_{11}}\right) - \frac{1}{2}\right)^+, \tag{15}$$

$$\Theta_{1,1} + \Theta_{2,3} + \Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1 + h_{21}\right) - \frac{1}{2}\right)^+, \tag{16}$$

$$\Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1 + h_{22}\right) - \frac{1}{2}\right)^+, \tag{17}$$

$$\Theta_{1,1} + \Theta_{2,1} + \Theta_{2,2} + \Theta_{2,3} + \Theta_{2,4} + \Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1 + h_{21} + h_{22}\right) - \frac{1}{2}\right)^+, \tag{18}$$

is achievable.

V.2 The Gaussian ZZ *network: Achievability*

The encoding scheme needed for the ZZ network is slightly more sophisticated than the ZS network. An additional component to strategic message splitting is that of interference neutralization. This was illustrated in examples 2 and 3 in Section III. This along with message splitting inspired by the network decomposition illustrated in example 1 of Section III, form the basis of the encoding scheme for the ZZ network.

More formally, the interference that has to be neutralized, will be combined with the main message in the first layer according to some partial-invertible function. In the second layer the inverse of the function is applied on this combination and the other interference received through the cross link. The remaining parts of the interference has to be either decoded or treated as noise. The neutralization is implemented using lattice codes and the rate-splitting along with appropriate power allocation is also used.

We formally define a partial-invertible function and a Z-*neutralization* network in the following. The Gaussian ZZ network is essentially a cascade of two Z-neutralization networks. An achievable rate region for the Z-*neutralization* network is given in Lemma 3. This rate region will be later used to obtain an achievable rate region for the Gaussian ZZ network. We will analyze the performance of the Gaussian encoding/decoding schemes in Appendix B.2.

Definition 1: Let $\mathcal{U}$ and $\mathcal{V}$ be two finite sets. A function $\phi(\cdot,\cdot)$ defined on $\mathcal{U}\times\mathcal{V}$ is called *partial-invertible*, if and only if having $\phi(u,v)$ and u, one can always reconstruct v for any $u\in\mathcal{U}$ and $v\in\mathcal{V}$. Similarly, u can be obtained from $\phi(u,v)$ and v.

An intuitive way of thinking about a partial-invertible $\phi(u,v)$ is the following. An arbitrary function defined on a finite sets $\mathcal{U}$ and $\mathcal{V}$ creates a table with rows corresponding to the elements of $\mathcal{U}$ and columns corresponding to the elements of $\mathcal{V}$, the each cell of the table consists the value assigned to its row and column by the function. A function will be partial-invertible, if and only if no two cells in the same column or row of its table be identical.

Note that summation over real numbers, and multiplication over non-zero numbers are two examples of partial-invertible functions. However, it is clear multiplication over real numbers is not partial-invertible, since $w=\phi(1,0)=\phi(2,0)$, and therefore having w and $v=0$, u can be anything.

Definition 2: Consider the Z network shown in Fig 12, which consists of a Gaussian broadcast channel from F_2 to the receivers and a Gaussian multiple access channel from F_1 and F_2 to G_1. A Z-*neutralization*

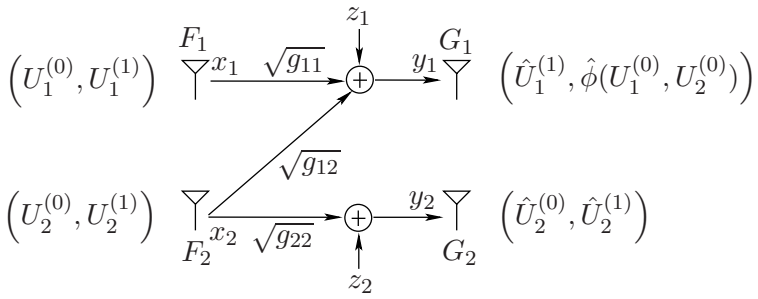


Fig. 12. The Gaussian Z channel.

network is a Z network, wherein the first source node has two messages $(U_1^{(0)},U_1^{(1)})$ of rates Υ_0 and Υ_1, respectively. Similarly the second source observes two independent messages $(U_2^{(0)},U_2^{(1)})$ of rates Υ_0 and Υ_2.

The second receiver is interested in decoding $U_2^{(0)}$ and $U_2^{(1)}$, while the first destination wishes to decode $\phi(U_1^{(0)},U_2^{(0)})$ and $U_1^{(1)}$, where $\phi(\cdot,\cdot)$ can be any arbitrary partial-invertible function. A rate tuple $(\Upsilon_0,\Upsilon_1,\Upsilon_2)$ is called achievable if the receivers can decode their messages with arbitrary small error probability.

***Lemma** 3:* Consider the Z-neutralization network defined Definition 2 with channel gains (g_{11},g_{12},g_{22}) (see Figure 12). Let

$$\lambda\triangleq\min\{g_{11},g_{12},g_{22}\},\tag{19}$$

and

$$\mu \triangleq \max\left\{g_{11}, g_{12}, g_{22}, \frac{g_{11}g_{22}}{g_{12}}\right\}. \tag{20}$$

Any rate tuple $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ satisfying

$$\Upsilon_0 \leq \left(\frac{1}{2}\log(\lambda) - \frac{1}{2}\right)^+, \tag{21}$$

$$\Upsilon_0 + \Upsilon_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \tag{22}$$

$$\Upsilon_0 + \Upsilon_2 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \tag{23}$$

$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \left(\frac{1}{2}\log(\mu) - \frac{3}{2}\right)^+, \tag{24}$$

is achievable.

As mentioned before, we strategically split the messages and require functional reconstructions for some of them at the relay nodes to facilitate neutralization at the destinations. More precisely, in the first layer of the network, each source node splits its message into two parts, namely, "functional" and private parts, $W_1 = \left(U_1^{(0)}, U_1^{(1)}\right)$ and $W_2 = \left(U_2^{(0)}, U_2^{(1)}\right)$. The "functional" parts $U_1^{(0)}, U_2^{(0)}$ both have the same rates Υ_0. Both transmitters use a common lattice code to encode their functional sub-messages. Now the first layer encodes the message such that the first receiver (which is relay A in the original ZZ network) can decode $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$, and the second one (relay B in the original ZZ network) can decode $U_2^{(0)}$ and $U_2^{(1)}$. Lemma 3 gives the rates at which these can be sent reliably. The second stage operates in a manner similar to the first stage, by splitting the messages into functional and private parts. The first sender (relay A in the original network) uses $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$ as the private and functional parts and the other one (relay B) uses $U_2^{(1)}$ and $U_2^{(0)}$ as the private and functional parts.

The functional parts are sent appropriately, using a common lattice code in both stages. Let $\mathbf{x}_1^{(N)}$ and $\mathbf{x}_2^{(N)}$ be the lattice codewords, corresponding to $U_1^{(0)}$ and $U_2^{(0)}$, respectively. The power allocation In the first layer it is done so that two lattice points get received at A at the same power (see Figure 8). The group structure of the lattice code implies that the summation of two received lattice point, $\tilde{\mathbf{x}}^{(N)} = \mathbf{x}_1^{(N)} + \mathbf{x}_2^{(N)}$ is still a valid codeword, and can be decoded by A. The function $\phi(\cdot,\cdot)$ is in fact the decoded message from $\tilde{\mathbf{x}}^{(N)}$. In the second stage, relay node B, sends the inverse of the the received lattice point, that is $\mathbf{x}_2^{'(N)} = -\mathbf{x}_2^{(N)}$, while A forwards the sum lattice point, $\mathbf{x}_1^{'(N)} = \tilde{\mathbf{x}}^{(N)}$. Again these lattice points are scaled properly so that they get received at D_1 at the same power. Thus, their summation would be a lattice point and equals $\mathbf{x}_1^{'(N)} + \mathbf{x}_2^{'(N)} = (\mathbf{x}_1^{(N)} + \mathbf{x}_2^{(N)}) - \mathbf{x}_2^{(N)} = \mathbf{x}_1^{(N)}$, which will be decoded to $U_1^{(0)}$.

The other destination D_2, receives $-\mathbf{x}_2^{(N)}$, finds its inverse $\mathbf{x}_2^{(N)}$, and finally decodes it to $U_2^{(0)}$. This idea is illustrated in Example 3, and the precise details of this argument are given in Appendix B.2.

VI. The Deterministic ZS Network

In this section we prove Theorem 1. We study this problem in two parts. First we present the converse proof, which shows any achievable rate pair belongs to $\mathcal{R}^{\mathsf{DZS}}$. Then for any rate pair in this region, we propose an encoding scheme which is able to transmit messages up to the desired rates.

VI.1 The Outer Bound

In this section we show that any achievable rate pair (R_1, R_2) for the deterministic ZS network belongs to $\mathcal{R}^{\mathsf{DZS}}$. Assume there exists a coding scheme with block length ℓ which can be used to communicate at rates R_1 and R_2 over the network. We use fold face matrices to denote ℓ copy of them, as the transfer matrix applied over a codeword of length ℓ, *e.g.,* $\mathbf{M}_{11} = I_\ell \otimes M_{11}$.

All of the bounds in the theorem except (DZS-3) and (DZS-10) can be obtained straight-forwardly using the generalized cut-set bound in [16], which shows that in a linear finite-field network, the maximum reliable rate can be transmitted through a cut is upper bounded by the rank of the transition matrix of the cut. Here, we only present the proof of (DZS-5) to illustrate this idea. Then we prove the two remaining bounds, which are tighter than the cut-set bound.

(DZS-5) $R_1 + R_2 \leq m_{22} + \max(n_{11}, n_{21})$*:* This bound corresponds to the cut $\Omega_s = \{S_1, S_2, A\}$ and $\Omega_d = \{B, D_1, D_2\}$. The transition matrix from the input of the cut $X_{\Omega_s} = (X_2, X_1')$ to its output $Y_{\Omega_d} = (Y_2', Y_1, Y_2)$ can be written as

$$\begin{bmatrix} Y_2^{'\ell} \\ Y_1^{\ell} \\ Y_2^{\ell} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{M}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{11} \\ \mathbf{0} & \mathbf{N}_{21} \end{bmatrix}}_{\mathbf{G}_{\Omega_s,\Omega_d}} \begin{bmatrix} X_2^{\ell} \\ X_1^{'\ell} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{N}_{22} \end{bmatrix} X_2^{'\ell}. \tag{25}$$

Therefore, from [16] we have

$$\ell(R_1 + R_2) \leq \operatorname{rank}(\mathbf{G}_{\Omega_s,\Omega_d}) = \operatorname{rank}(\mathbf{M}_{22}) + \operatorname{rank}\left(\begin{bmatrix} \mathbf{N}_{11} \\ \mathbf{N}_{21} \end{bmatrix}\right) = \ell m_{22} + \ell \max(n_{11}, n_{21}). \tag{26}$$

As mentioned before, we skip the proof of those bounds which follow from the generalized cut-set bound. In the following we present the proof of the two remaining inequalities which are tighter that the cut-set bound.

(DZS-3) $R_1 + R_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+$*:* In order to prove this bound, we can start with

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(X_1^\ell, X_2^\ell; Y_1^\ell, Y_2^\ell) \\
&\leq I(X_1^\ell, X_2^\ell; Y_1^{'\ell}, Y_2^{'\ell}) \qquad (27) \\
&= I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) + I(X_1^\ell, X_2^\ell; Y_2^{'\ell}|Y_1^{'\ell}) \\
&\leq I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) + H(Y_2^{'\ell}|Y_1^{'\ell}) - H(Y_2^{'\ell}|X_1^\ell, X_2^\ell, Y_1^{'\ell}), \qquad (28)
\end{aligned}$$

where in (27) we used the data-processing inequality for the Markov chain

$$(X_1^\ell, X_2^\ell) \leftrightarrow (Y_1^{'\ell}, Y_2^{'\ell}) \leftrightarrow (X_1^{'\ell}, X_2^{'\ell}) \leftrightarrow (Y_1^\ell, Y_2^\ell), \qquad (29)$$

and (28) holds since $Y_2^{'\ell}$ is function of X_2^ℓ. Now, it is clear that

$$I(X_1^\ell, X_2^\ell; Y_1^\ell) \leq \text{rank}\left(\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{21} \end{bmatrix}\right) = \ell \max(m_{11}, m_{12}). \qquad (30)$$

In order to bound the second term, we can write

$$\begin{aligned}
H(Y_2^{'\ell}|Y_1^{'\ell}) &= H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell) \qquad (31) \\
&\leq H(Y_2^{'\ell}, W_2|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell) \\
&= H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell, W_2) + H(W_2|Y'\ell_1, X_1^{'\ell}, Y_1^\ell) \\
&\leq H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell, W_2, X_1^\ell) + \ell\varepsilon_\ell \qquad (32) \\
&\leq H(Y_2^{'\ell}|Y_1^{'\ell} - \mathbf{M}_{11}X_1^\ell) + \ell\varepsilon_\ell \\
&\leq H(M_{22}X_2^\ell|\mathbf{M}_{12}X_2^\ell) + \ell\varepsilon_\ell \\
&\leq \ell\text{rank}\left(\begin{bmatrix} M_{12} \\ M_{22} \end{bmatrix}\right) - \ell\text{rank}\,(M_{12}) + \ell\varepsilon_\ell \\
&= \ell(m_{22} - m_{12})^+ + \ell\varepsilon_\ell, \qquad (33)
\end{aligned}$$

where (31) holds since $X_1^{'\ell}$ is also a function of $Y_1^{'\ell}$, and Y_1^ℓ is also a deterministic function of $X_1^{'\ell}$.. We used Fano's inequality in (32), where W_1 should be decodable based on Y_1^ℓ. Summing up (30) and (33), we get the desired bound.

Note that the cut-set bound for the cut $\Omega_s = \{S_1, S_2\}$ and $\Omega_d = \{A, B, D_1, D_2\}$ gives us

$$\ell(R_1 + R_2) \leq \text{rank}\left(\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{0} & \mathbf{M}_{22} \end{bmatrix}\right) = \ell \max(m_{11} + m_{22}, m_{12}), \qquad (34)$$

in which the RHS can be arbitrarily larger than the RHS of the presented bound. The reason for this difference is the following. It is inherently assumed in deriving the cut-set bound that the receivers can cooperate to decode the messages of rates R_1 and R_2, and no decodability requirement is posed for individual receivers. However, the setup of this problem impose an extra constraint, that is B alone should be able to decode W_2. Incorporating this decodability requirement shrinks the set of admissible rates, and gives us a tighter bound.

(DZS-10) $R_1 + R_2 \leq \max(n_{21}, n_{11}) + (n_{11} - n_{21})^+$*:* The last inequality captures the maximum flow of information from the relays to the destinations, such that D_1 and D_2 be able to decode W_1 and W_2, respectively. We again start with

$$\ell(R_1 + R_2) \leq I(X_1^\ell, X_2^\ell; Y_1^\ell, Y_2^\ell) = H(Y_1^\ell, Y_2^\ell) = H(Y_2^\ell) + H(Y_1^\ell | Y_2^\ell). \tag{35}$$

The first term can be easily bounded by

$$H(Y_2^\ell) \leq \text{rank}\left(\left[\begin{array}{cc} \mathbf{N}_{21} & \mathbf{N}_{22} \end{array}\right]\right) = \ell \max(n_{21}, n_{22}). \tag{36}$$

In order to bound the second term, we use the fact that W_2 can be decoded from Y_2^ℓ. Therefore,

$$\begin{aligned}
H(Y_1^\ell | Y_2^\ell) &\leq H(Y_1^\ell, W_2 | Y_2^\ell) \\
&= H(Y_1^\ell | Y_2^\ell, W_2) + H(W_2 | Y_2^\ell) \\
&\leq H(Y_1^\ell | Y_2^\ell, W_2) + \ell\varepsilon_\ell \qquad (37) \\
&= H(Y_1^\ell | Y_2^\ell, W_2, X_2^\ell, Y_2^{'\ell}, X_2^{'\ell}) + \ell\varepsilon_\ell \\
&\leq H(Y_1^\ell | Y_2^\ell - \mathbf{N}_{22} X_2^{'\ell}) + \ell\varepsilon_\ell \\
&= H(N_{11} X_1^{'\ell} | \mathbf{N}_{21} X_1^{'\ell}) + \ell\varepsilon_\ell \\
&\leq \ell \text{rank}\left(\left[\begin{array}{c} N_{11} \\ N_{21} \end{array}\right]\right) - \ell \text{rank}\,(N_{21}) + \ell\varepsilon_\ell \\
&= \ell(n_{11} - n_{21})^+ + \ell\varepsilon_\ell. \qquad (38)
\end{aligned}$$

In (37) we used the Fano's inequality, as well as the fact that X_2^ℓ, $Y_2^{'\ell}$, and $X_2^{'\ell}$ are known having W_2. The bound is obtained by replacing (36) and (38) in (35).

It is worth mentioning that this bound is tighter than the cut-set bound for the cut $\Omega_s = \{S_1, S_2, A, B\}$ and $\Omega_d = \{D_1, D_2\}$, which is

$$R_1 + R_2 \leq \max(n_{11} + n_{22}, n_{12}). \tag{39}$$

VI.2 The Achievability Part

Network Decomposition:: The achievability scheme presented here is based on decomposition of the deterministic ZS network into two node-disjoint networks. In fact, such partitioning depends on the demanded rate pair $(R_1, R_2) \in \mathcal{R}^{\mathsf{DZS}}$. The resulting family of separations immediately suggests a simple coding scheme. We will show that this separation is optimal, and does not cause any loss in the admissible rate region of the network.

Before introducing the network decomposition, we define an equivalence class for the sub-nodes (levels) in a network.

Definition 3: In a Z (or S) deterministic network, two sub-nodes a and b are called *related* sub-nodes, and denoted by $a \sim b$ if any of the following conditions hold:

- $a = b$;
- a is connected to b;
- b is connected to a;
- there exists a sub-node c such that c broadcasts to both a and b;
- there exists a sub-node d where both a and b are connected to.

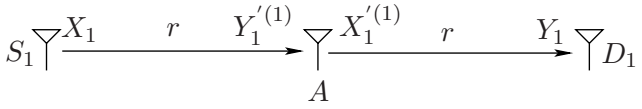
Note that this relation is reflective, symmetric, and transitive. Therefore, it forms equivalence classes for the sub-nodes.

We denote by $\mathcal{N}_1$ and $\mathcal{N}_2$ the partitions of the network. Assume we wish transmitting at rate $R_1 = r \leq \min(m_{11}, n_{11})$ from S_1 to D_1. The first part of the network $\mathcal{N}_1$, includes the top $(m_{11} - m_{12})^+$ levels as well as the lowest $(r - (m_{11} - m_{12})^+)^+$ levels of S_1. It also includes all the related sub-nodes of S_2, and the receiver levels of A and B. Similarly, in the second layer of the network, $\mathcal{N}_1$ includes the lowest $(n_{11} - n_{21})^+$ levels as well as the top $(r - (n_{11} - n_{21})^+)^+$ nodes of the transmitter part of A. All related sub-nodes of the transmitter part of B, as well as D_1 and D_2 also belong to $\mathcal{N}_1$. The second part of the network $\mathcal{N}_2$, is formed by all the remaining nodes.

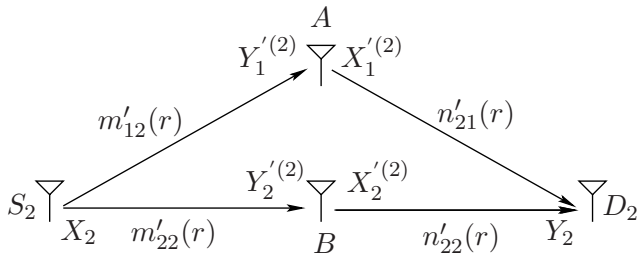
We will use $\mathcal{N}_1$ for transmitting data from S_1 to D_1. Similarly $\mathcal{N}_2$ is only used to communicate from S_2 to D_2. Therefore, we have two uni-cast networks, and each pair of transmitter-receiver can communicate up to the capacity of their own partition, which is the min-cut of the partition [4].

It is worth mentioning that any two "related" sub-nodes belong to the same partition. Therefore, these two networks are node-disjoint, and do not cause interference for each other. This allows us to derive the capacity of each network separately, and argue that (R_1, R_2) can be achieved simultaneously for the original network, if R_1 and R_2 are achievable for partitions $\mathcal{N}_1$ and $\mathcal{N}_2$.

Encoding Scheme: A transmission from S_1 and S_2 to D_1 and D_2 is performed as follows. S_1 transmits only on its sub-nodes which belong to $\mathcal{N}_1$, and keeps its other sub-nodes silent. Similarly, S_2 encodes its message on the sub-nodes included in $\mathcal{N}_2$, and sends zero on the other levels. Therefore, the *effective* communication over each partition is a simple uni-cast.



(a) Effective channel for (S_1, D_1).



(b) Effective channel for (S_2, D_2).

Fig. 13. The effective separated ZS network.

Fig. 13 shows the effective parts of the network. It is easy to see that the diamond network in Figure 13(b) is also a linear shift deterministic networks, with channel gains

$$m'_{12}(r) = \min(\max(m_{11}, m_{12}) - r, m_{12}), \tag{40}$$

$$m'_{22}(r) = \min(\max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ - r, m_{22}), \tag{41}$$

$$n'_{21}(r) = \min(\max(n_{11}, n_{21}) - r, n_{21}), \tag{42}$$

$$n'_{22}(r) = \min(\max(n_{11}, n_{21}) + (n_{22} - n_{21})^+ - r, n_{22}). \tag{43}$$

Achievable Rate Region: The cut values of $\mathcal{N}_1$ can be easily computed as

$$\Omega = \{S_1\}: \qquad (m_{11} - m_{12})^+ + (r - (m_{11} - m_{12})^+)^+ = \max\{(m_{11} - m_{12})^+, r\} \geq r$$

$$\Omega = \{S_1, A\}: \qquad (r - (n_{11} - n_{12})^+)^+ + (n_{11} - n_{12})^+ = \max\{(n_{11} - n_{12})^+, r\} \geq r.$$

Therefore any rate in $\mathcal{R}_1^{\mathsf{DZS}}(r) = \{R_1 : R_1 \leq r\}$ can be conveyed from S_1 to D_1 through $\mathcal{N}_1$.

The capacity of $\mathcal{N}_2$ can be found using the generalized max-flow min-cut theorem [4]. Hence, the rate

region of the second partition $\mathcal{N}_2$ would be

$$\mathcal{R}_2^{\mathsf{DZS}}(r) = \{R_2 : R_2 \leq \max(m'_{12}(r), m'_{22}(r)), \tag{44}$$

$$R_2 \leq m'_{22}(r) + n'_{21}(r), \tag{45}$$

$$R_2 \leq m'_{12}(r) + n'_{22}(r), \tag{46}$$

$$R_2 \leq \max(n'_{21}(r), n'_{22}(r))\}. \tag{47}$$

Therefore, by using this decomposition, any rate pair in the set $\mathcal{R}_1^{\mathsf{DZS}}(r) \times \mathcal{R}_2^{\mathsf{DZS}}(r) = \{(R_1, R_2) : R_1 \in \mathcal{R}_1^{\mathsf{DZS}}(r), R_2 \in \mathcal{R}_2^{\mathsf{DZS}}(r)\}$ can be achieved. It remains to prove the following lemma.

***Lemma** 4:* For any deterministic ZS network,

$$\mathcal{R}^{\mathsf{DZS}} \subseteq \bigcup_{r \leq \min(m_{11}, n_{11})} \left(\mathcal{R}_1^{\mathsf{DZS}}(r) \times \mathcal{R}_2^{\mathsf{DZS}}(r) \right). \tag{48}$$

We will prove this lemma in Appendix C.

VII. The Deterministic ZZ Network

In this section we prove Theorem 3. This is done in two parts, that provide the converse and achievability proofs.

VII.1 The Outer Bound

In the following we will show that any achievable rate pair (R_1, R_2) satisfies constraints (DZZ-1)-(DZZ-6). The individual rate bounds can be directly obtained by the generalized cut-set bound introduced in [16], where the maximum flow of information through a cut in a linear deterministic network is upper bounded by the rank of the transition matrix from the sender part of the cut to its receiver part. Hence, we skip the proofs of (DZZ-1)-(DZZ-4).

The sum-rate bounds in (DZZ-5)-(DZZ-6) are, however, genie-aided bounds which are tighter that the cut-set bounds. In the following, we focus on these two bounds, and present their proofs in detail. Again we assume that there exists a coding scheme with block length ℓ which can be used to communicate at rates R_1 and R_2 over the network.

(DZZ-5) $R_1 + R_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ + n_{12}$*:* In order to prove this inequality we focus on the flow of information from the sources to the relays. The key idea here is to provide A with the information sent by B to D_1 as side information. In such condition, the information A has received about W_1 is stronger than the information available at D_1, and therefore A can decode W_1 since D_1 can

as well. Once W_1 is decoded at A, it can determine the transmitted codeword from S_1. By removing the interference from S_1, A can also partially decode W_2.

More precisely, we can write

$$\begin{aligned}\ell(R_1+R_2) &\le I(X_1^\ell, X_2^\ell; Y_1^{'\ell}, Y_2^{'\ell}) = H(Y_1^{'\ell}, Y_2^{'\ell}) \le H(Y_1^{'\ell}, Y_2^{'\ell}, \Gamma_2^\ell) \\ &= H(Y_1^{'\ell}, \Gamma_2^\ell) + H(Y_2^{'\ell}|Y_1^{'\ell}, \Gamma_2^\ell) \\ &\le H(Y_1^{'\ell}) + H(\Gamma_2^\ell) + H(Y_2^{'\ell}|Y_1^{'\ell}, \Gamma_2^\ell), \end{aligned} \tag{49}$$

where $\Gamma_2^\ell = \mathbf{N}_{12} X_2^{'\ell}$ is the part of the signal received at D_2 from B as in Figure 4(b). The first two terms are easily bounded by $\ell \max(m_{11}, m_{12})$ and ℓn_{12}, respectively. Deriving an upper bound for the last term is more involved.

Similar to Γl_2, we define $\Gamma_1^\ell = \mathbf{M}_{12} X_2^\ell$, where we have

$$\begin{aligned} H(\Gamma_1^\ell|Y_1^{'\ell}, \Gamma l_2) &= H(Y_1^{'\ell} - \mathbf{M}_{11} X_1^\ell | Y_1^{'\ell}, \Gamma l_2) \qquad (50) \\ &\le H(X_1^\ell|Y_1^{'\ell}, \Gamma l_2) \\ &\le H(W_1|Y_1^{'\ell}, \Gamma l_2) \\ &= H(W_1|Y_1^{'\ell}, X_1^{'\ell}, \Gamma l_2) \\ &\le H(W_1|\mathbf{N}_{11} X_1^{'\ell} + \Gamma l_2) \\ &= H(W_1|Y l_1) \le \ell \varepsilon_\ell, \qquad (51) \end{aligned}$$

where $\varepsilon_\ell \to 0$ as ℓ grows. We have used the invertibility property of the deterministic multiple access channel in (50), and (51) follows from the Fano's inequality, and the fact that D_1 can decode the message sent by S_1. Therefore, we have $H(\Gamma l_1|Y_1^{'\ell}, \Gamma l_2) \le \ell \varepsilon_\ell$. Hence,

$$\begin{aligned} H(Y_2^{'\ell}|Y_1^{'\ell}, \Gamma l_2) &\le H(Y_2^{'\ell}, \Gamma l_1|Y_1^{'\ell}, \Gamma l_2) \\ &= H(Y_2^{'\ell}|\Gamma l_1, Y_1^{'\ell}, \Gamma_2) + H(\Gamma l_1|Y_1^{'\ell}, \Gamma l_2) \\ &\le H(Y_2^{'\ell}|\Gamma l_1) + \ell \varepsilon_\ell \\ &= H(\mathbf{M}_{22} X_2^\ell | \mathbf{M}_{12} X_2^\ell) + \ell \varepsilon_\ell \\ &\le \ell (m_{22} - m_{12})^+ + \ell \varepsilon_\ell. \qquad (52) \end{aligned}$$

Replacing the upper bounds for each term in (49), we get

$$R_1 + R_2 \le \max(m_{11}, m_{12}) + n_{12} + (m_{22} - m_{12})^+. \tag{53}$$

It is worth mentioning that the cut-set bound for $\Omega_s = \{S_1, S_2\}$ and $\Omega_d = \{A, B, D_1, D_2\}$ gives us

$$R_1 + R_2 \leq \max(m_{11} + m_{22}, m_{12}), \tag{54}$$

which is looser than the genie-aided bound.

(DZZ-6) $R_1+R_2 \leq \max(n_{11}, n_{12})+(n_{22}-n_{12})^+ + m_{12}$*:* The last inequality captures the maximum flow of information from the relays to the destinations. Intuitively, this inequality says that the number of interfering bits can get neutralized at D_1 cannot exceed the minimum of m_{12} and n_{12}. In order to make this intuition formal, we provide Γl_1, the partial information about W_2 which is available at A, as side information for D_1. We then have

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(Yl_1, Yl_2; Xl_1, Xl_2) = H(Yl_1, Yl_2) \\
&\leq H(Yl_1, Yl_2, \Gamma l_1) \\
&\leq H(Yl_1) + H(\Gamma l_1) + H(Yl_2|Yl_1, \Gamma l_1).
\end{aligned}$$

Again, we can simply upper bound the first two terms by the rank of the corresponding matrices. In order to bound the last term, similar to the proof of (DZZ-5), we use the following bounding technique.

$$\begin{aligned}
H(\Gamma l_2|Yl_1, \Gamma_1^\ell) &= H(Yl_1 - \mathbf{N}_{11}X_1^{'\ell}|Yl_1, \Gamma l_1) \\
&\leq H(X_1^{'\ell}|Yl_1, \Gamma l_1) \\
&\leq H(Y_1^{'\ell}|Yl_1, \Gamma l_1) \\
&= H(\mathbf{M}_{11}Xl_1 + \Gamma l_1|Yl_1, \Gamma l_1) \\
&\leq H(Xl_1|Yl_1, \Gamma l_1) \\
&\leq H(Xl_1|Yl_1) \\
&\leq H(W_1|Yl_1) \leq \ell\varepsilon_\ell,
\end{aligned} \tag{55}$$

where (55) follows from the Fano's inequality. This inequality can be used as

$$\begin{aligned}
H(Yl_2|Yl_1, \Gamma l_1) &\leq H(Yl_2, \Gamma l_2|Yl_1, \Gamma l_1) \\
&= H(Yl_2|\Gamma l_2, Yl_1, \Gamma l_1) + H(\Gamma l_2|Yl_1, \Gamma l_1) \\
&\leq H(Yl_2|\Gamma l_2) + \ell\varepsilon_\ell \\
&= H(\mathbf{N}_{22}X_2^{'\ell}|\mathbf{N}_{12}X_2^{'\ell}) + \ell\varepsilon_\ell \\
&\leq \ell(n_{22} - n_{12})^+ + \ell\varepsilon_\ell.
\end{aligned} \tag{56}$$

Therefore, we have

$$R_1 + R_2 \leq \max(n_{11}, n_{12}) + m_{12} + (n_{22} - n_{12})^+. \tag{57}$$

Again, it is easy to show that this bound is tighter than the cut-set bound for $\Omega_s = \{S_1, S_2, A, B\}$ and $\Omega_d = \{D_1, D_2\}$,

$$R_1 + R_2 \leq \max(n_{11} + n_{22}, n_{12}). \tag{58}$$

This completes the proof of the converse part of Theorem 3.

VII.2 The Achievability Proof

In this part we will show that all rate pairs satisfying inequalities (DZZ-1)-(DZZ-6) are achievable. In particular, we introduce a coding scheme which achieves such rates. Our coding strategy provides the interference neutralization at the destination. This is performed by splitting the messages into two parts, namely *private* and *functional* parts. The private sub-messages can be decoded at the relays, and forwarded to the destinations. The functional sub-message of the second source can be also decoded at B. However, A only receives a combination (xor) of the functional sub-messages, and cannot decode them. It only forwards such combination on proper (power) levels such that the interference caused by the functional sub-message of S_2 get neutralized over the second layer of the network, and D_1 can decode the sub-message of its interest.

Our analysis is based on characterizing the number of *pure* and *combined* bits can be sent through each layer of the network. In the following we focus on one layer of the network, and obtain an achievable rate region for these numbers. Next, we use this region to build the encoding scheme for the ZZ network, and obtain an achievable rate region, which matches with the outer bound.

Definition 4: Consider a deterministic Z network, with gains (n_{11}, n_{12}, n_{22}). as shown in Figure 14. Each of the transmitters has a set of information bits to transmit to the receivers. This set for F_i includes Υ_i private bits and Υ_0 functional bits, namely, $\mathcal{W}_{i,P} = \{W_{i,P}(1), \ldots, W_{i,P}(\Upsilon_i)\}$ and $\mathcal{W}_{i,N} = \{W_{i,N}(1), \ldots, X_{i,N}(\Upsilon_0)\}$. The second receiver wishes to receive all the private and functional bits of F_2, while the first receiver is interested in receiving the private bits of F_1, and the xor of the functional bits of F_1 and F_2. More precisely, denoting by $\hat{\mathcal{W}}_i$ the set of bits G_i is interested in, we have

$$\hat{\mathcal{W}}_1 = \mathcal{W}_{1,P} \cup \{\tilde{W}_{1,N}(j) \triangleq W_{1,N}(j) \oplus W_{2,N}(j) : j = 1, \ldots, \Upsilon_0\},$$
$$\hat{\mathcal{W}}_2 = \mathcal{W}_{2,P} \cup \mathcal{W}_{2,N}.$$

We term this network with the described decoding demands as deterministic Z-neutralization network. The goal is the characterize the set achievable tuples $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$.

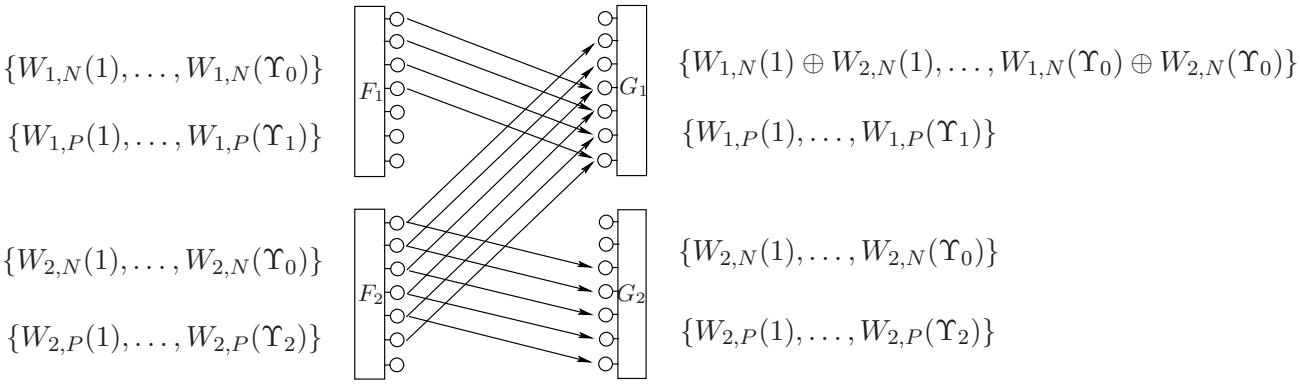


Fig. 14. A deterministic Z-neutralization network with the message demands.

The following lemma gives an achievable rate region for the deterministic Z-neutralization network. The proof of this lemma can be found in Appendix C.

***Lemma** 5:* Consider the deterministic Z-neutralization network defined in Definition 4 with channel gains (n_{11}, n_{12}, n_{22}) (see Figure 14). Any rate tuple $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ satisfying

$$\Upsilon_0 \leq \lambda \triangleq \min\{n_{11}, n_{12}, n_{22}\}, \tag{59}$$

$$\Upsilon_0 + \Upsilon_1 \leq n_{11}, \tag{60}$$

$$\Upsilon_0 + \Upsilon_2 \leq n_{22}, \tag{61}$$

$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \mu \triangleq \max\{n_{11}, n_{12}, n_{22}, n_{11} + n_{22} - n_{12}\}. \tag{62}$$

is achievable for this network.

Now, having an achievable rate region for the deterministic Z-neutralization network, we are ready to present the coding scheme and analyze its rate region for the ZZ network.

Recall that the ZZ network consists of two cascaded Z network. In first layer, the source nodes split their message into private and functional parts. They can send these parts to the relays as long as their rates belong to the achievable rate region of the first layer given in Lemma 5. Once the relays receive these sub-messages, forward them to the destination nodes using the same scheme for the private and functional sub-messages. This can be done if the rate tuple for the sub-messages satisfy the corresponding inequalities for the second layer as well. Note that functional bits received at the destination are $\tilde{W}_{1,N}(j) \oplus W_{2,N}(j) = [W_{1,N}(j) \oplus W_{2,N}(j)] \oplus W_{2,N}(j) = W_{1,N}(j)$. Therefore, the interference of these bits get neutralized, and pure information bits will be received at the destination.

The achievable rate region of this scheme is given by

$$\begin{aligned}\mathcal{R}_{\text{ach}}^{\mathsf{DZZ}} = \Big\{(R_1,R_2) :& \exists \Upsilon_0,\Upsilon_1,\Upsilon_2 \geq 0,\\ & R_1 = \Upsilon_0 + \Upsilon_1,\\ & R_2 = \Upsilon_0 + \Upsilon_2,\\ & \Upsilon_0 \leq \min\{\lambda_m,\lambda_n\},\\ & \Upsilon_0 + \Upsilon_1 \leq \min\{m_{11},n_{11}\},\\ & \Upsilon_0 + \Upsilon_2 \leq \min\{m_{22},n_{22}\},\\ & \Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \min\{\mu_m,\mu_n\}\Big\} \end{aligned} \tag{63}$$

Here we used subscripts m and n to denote λ and μ parameters of the first and the second layer of the network, respectively. Applying Fourier-Motzkin elimination on this set to project it on the (R_1,R_2) plane, gives us the rate region claimed in the theorem.

VIII. Discussion

Interference management is perhaps the most fundamental open problem in wireless networks. The recent progress in (approximate) characterization of the interference channel capacity and the utility of the deterministic approach inspired the questions studied in this paper. Even though the interference-relay networks studied in this work were special, they revealed several new features needed for information transmission. In particular, the interference neutralization and network flow decomposition techniques were uncovered through the study of ZZ and ZS networks. We also saw the importance of using structured lattice codes for interference neutralization. Moreover, we believe that the neutralization technique is robust to channel uncertainties and one could get partial neutralization in such situations. This is a topic of ongoing work on this topic. We also believe that the outer bounding techniques developed in this work could have more general applicability in the wireless multiple-unicast problem. The two-unicast problem in arbitrary layered wireless networks would be a natural next step arising out of our work. The deterministic approach for this problem has already provided some interesting new techniques [17]. In summary we believe that the deterministic approach is a promising methodology to make progress on the wireless multiple-unicast problem.

APPENDIX A
THE GAUSSIAN ZS NETWORK

A.1 The Outer Bound

In the following we will prove each of the inequalities in (GZS-1)-(GZS-10), separately. We will use the notation as shown in Figure 15, and assume that the rate pair (R_1, R_2) can be achieved with small enough decoding error probability using a code of length ℓ.

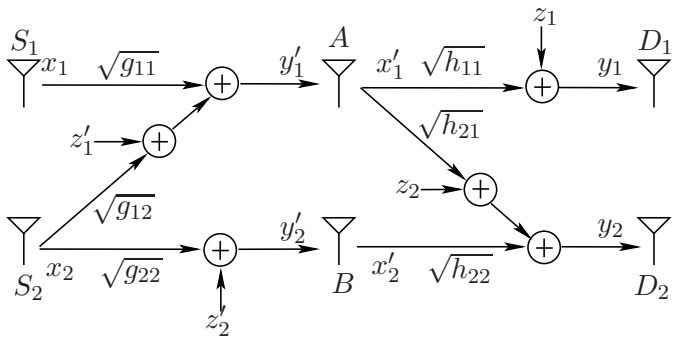


Fig. 15. The Gaussian ZS network.

***Lemma** 6:* Any achievable rate pair (R_1, R_2) satisfies

$$\ell R_1 \leq I(x_1^\ell; y_1^\ell) + \ell\varepsilon_\ell, \tag{A.1}$$

$$\ell R_2 \leq I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell, \tag{A.2}$$

$$\ell(R_1 + R_2) \leq I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\varepsilon_\ell. \tag{A.3}$$

Note that $\varepsilon_\ell \to 0$ as ℓ grows.

This lemma is a consequence of the Fano's lemma combined with the decodability requirements imposed by the problem, and its proof is given in Appendix C.

Most of the inequalities in (GZS-1)-(GZS-10) are cut-set type bounds, although the proof presented here are slightly different than the standard argument. However, the sum-rate bounds in (GZS-3) and (GZS-10) are different from the well known cut-set bounds. These two bounds are in general tighter than the cut values for the corresponding cuts. This is because the decoders are inherently allowed to cooperate in deriving a cut-set bound, while individual decoding abilities are imposed in this problem. In the following we first present the proofs of (GZS-3) and (GZS-10), which are more involved, and then prove the cut-set type bounds.

a) The proofs of non-cut-set type bounds:

- (GZS-3) $R_1+R_2 < \frac{1}{2}\log(1+g_{11}+g_{12})+\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right)$: We start with Lemma 6 for the sum-rate which implies

$$\begin{aligned}
\ell(R_1+R_2) &\le I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\varepsilon_\ell \\
&\le I(x_1^\ell, x_2^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\varepsilon_\ell && \text{(A.4)} \\
&= I(x_1^\ell, x_2^\ell; y_1^{'\ell}) + I(x_1^\ell, x_2^\ell; y_2^{'\ell}|y_1^{'\ell}) + \ell\varepsilon_\ell \\
&\le \frac{\ell}{2}\log(1+g_{11}+g_{12}) + h(y_2^{'\ell}|y_1^{'\ell}) - h(y_2^{'\ell}|y_1^{'\ell}, x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell, && \text{(A.5)}
\end{aligned}$$

where (A.4) follows from the data processing inequality. Now, note that

$$\begin{aligned}
h(y_2^{'\ell}, W_1|y_1^{'\ell}) &= h(y_2^{'\ell}|y_1^{'\ell}) + H(W_1|y_1^{'\ell}, y_2^{'\ell}) \\
&= H(W_1|y_1^{'\ell}) + h(y_2^{'\ell}|W_1, y_1^{'\ell}) \\
&\le H(W_1|y_1^\ell) + h(y_2^{'\ell}|W_1, y_1^{'\ell}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
h(y_2^{'\ell}|y_1^{'\ell}) &\le h(y_2^{'\ell}|W_1, y_1^{'\ell}) + \ell\varepsilon_\ell \\
&\le h(y_2^{'\ell}|x_1^\ell, y_1^{'\ell}) + \ell\varepsilon_\ell && \text{(A.6)} \\
&= h(y_2^{'\ell}|x_1^\ell, \sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\varepsilon_\ell && \text{(A.7)} \\
&\le h(\sqrt{g_{22}}x_2^\ell + z_2^{'\ell}|\sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\varepsilon_\ell \\
&= h(\sqrt{g_{22}}x_2^\ell + z_2^{'\ell} - \frac{\sqrt{g_{22}}}{\sqrt{g_{12}}}(\sqrt{g_{12}}x_2^\ell + z_1^{'\ell})|\sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\varepsilon_\ell \\
&\le h(z_2^{'\ell} - \frac{\sqrt{g_{22}}}{\sqrt{g_{12}}}z_1^{'\ell}) + \ell\varepsilon_\ell \\
&= \frac{\ell}{2}\log(2\pi e)\left(1+\frac{g_{22}}{g_{12}}\right) + \ell\varepsilon_\ell, && \text{(A.8)}
\end{aligned}$$

where (A.6) holds since x_1^ℓ is a function of W_1, and in (A.7) we used the invertibility property of the function $y_1^\ell = \sqrt{g_{11}}x_1^\ell + \sqrt{g_{12}}x_2^\ell + z_1^{'\ell}$. Replacing $h(y_2^{'\ell}|y_1^{'\ell})$ from (A.8) in (A.5), we get the desired bound.

- (GZS-10) $R_1+R_2 < \frac{1}{2}\log\left(1+\frac{h_{11}}{h_{21}}\right)+\frac{\ell}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}})$: The sum-rate can be

upper bounded as in Lemma 6. Next, we have

$$\begin{aligned}\ell(R_1+R_2) &\le I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\varepsilon_\ell \\ &\le I(x_1^{'\ell}, x_2^{'\ell}; y_1^\ell, y_2^\ell) + \ell\varepsilon_\ell \\ &= I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) + I(x_2^{'\ell}; y_1^\ell | y_2^\ell) + I(x_1^{'\ell}; y_1^\ell | x_2^{'\ell}, y_2^\ell) + \ell\varepsilon_\ell. \end{aligned} \tag{A.9}$$

The first term in (A.9) can be simply upper bounded as

$$I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) \le \frac{\ell}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}). \tag{A.10}$$

In order to bound the second term, we can use the fact that W_2 can be decoded from y_2^ℓ, and write

$$\begin{aligned} I(x_2^{'\ell}; y_1^\ell, W_2 | y_2^\ell) &= I(x_2^{'\ell}; y_1^\ell | y_2^\ell) + I(x_2^{'\ell}; W_2 | y_1^\ell, y_2^\ell) \\ &= I(x_2^{'\ell}; y_1^\ell | W_2, y_2^\ell) + I(x_2^{'\ell}; W_2 | y_2^\ell) \\ &\le I(x_2^{'\ell}; y_1^\ell | W_2, y_2^\ell) + H(W_2 | y_2^\ell) \\ &\le I(x_2^{'\ell}; y_1^\ell | y_2^\ell, W_2) + \ell\varepsilon_\ell. \end{aligned}$$

Therefore,

$$I(x_2^{'\ell}; y_1^\ell | y_2^\ell) \le I(x_2^{'\ell}; y_1^\ell | y_2^\ell, W_2) + \ell\varepsilon_\ell \le I(y_2^{'\ell}; y_1^\ell | y_2^\ell, W_2) + \ell\varepsilon_\ell = \ell\varepsilon_\ell, \tag{A.11}$$

where the second inequality follows from the fact that $x_2^{'\ell}$ is a function of $y_2^{'\ell}$, and (A.11) holds since $y_2^{'\ell}$ and y_1^ℓ are independent if W_2 is given.

Finally, we bound the last term as follows.

$$\begin{aligned} I(x_1^{'\ell}; y_1^\ell | x_2^{'\ell}, y_2^\ell) &= I(x_1^{'\ell}; y_1^\ell | x_2^{'\ell}, \sqrt{h_{21}}x_1^{'\ell} + z_2^\ell) \\ &= h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell | x_2^{'\ell}, \sqrt{h_{21}}x_1^{'\ell} + z_2^\ell) - h(y_1^\ell | x_2^{'\ell}, \sqrt{h_{21}}x_1^{'\ell} + z_2^\ell, x_1^{'\ell}) \\ &\le h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell - \frac{\sqrt{h_{11}}}{\sqrt{h_{21}}}(\sqrt{h_{21}}x_1^{'\ell} + z_2^\ell)) - h(z_1^\ell) \\ &\le \frac{\ell}{2}\log\left(1 + \frac{h_{11}}{h_{12}}\right). \end{aligned} \tag{A.12}$$

Replacing the bound derived for the three terms, (A.10), (A.11), and (A.12) in (A.9), we get the desired bound.

b) The proofs of cut-set type bounds:

- (GZS-1) $R_1 < \frac{1}{2}\log(1+g_{11})$: We start by Lemma 6, and write

$$\begin{aligned}
\ell R_1 &= I(x_1^\ell; y_1^\ell) + \ell\varepsilon_\ell \\
&\le I(x_1^\ell; y_1^{'\ell}) + \ell\varepsilon_\ell && \text{(A.13)} \\
&\le I(x_1^\ell; x_2^\ell, y_1^{'\ell}) + \ell\varepsilon_\ell \\
&= I(x_1^\ell; x_2^\ell) + I(x_1^\ell; y_1^{'\ell}|x_2^\ell) + \ell\varepsilon_\ell && \text{(A.14)} \\
&\le \frac{\ell}{2}\log(1+g_{11}) + \ell\varepsilon_\ell && \text{(A.15)}
\end{aligned}$$

where (A.13) follows from the data-processing inequality for the Markov chain $x_1^\ell \leftrightarrow y_1^{'\ell} \leftrightarrow x_1^{'\ell} \leftrightarrow y_1^\ell$, and in (A.14) we used the fact that x_1^ℓ and x_2^ℓ are independent. It is worth mentioning that this inequality essentially bounds the maximum flow that can be transmitted through the cut $\Omega_s = \{S_1\}$ and $\Omega_d = \{S_2, A, B, D_1, D_2\}$.

- (GZS-2) $R_2 < \frac{1}{2}\log(1+g_{12}+g_{22})$: Again starting from Lemma 6, we have

$$\begin{aligned}
\ell R_2 &\le I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell \\
&\le I(x_2^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\varepsilon_\ell && \text{(A.16)} \\
&\le I(x_2^\ell; x_1^\ell, y_1^{'\ell}, y_2^{'\ell}) + \ell\varepsilon_\ell \\
&= I(x_2^\ell; x_1^\ell) + I(x_2^\ell; y_1^{'\ell}, y_2^{'\ell}|x_1^\ell) + \ell\varepsilon_\ell \\
&= h(y_1^{'\ell}, y_2^{'\ell}|x_1^\ell) - h(y_1^{'\ell}, y_2^{'\ell}|x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\
&\le h(\sqrt{g_{12}}x_2^\ell + z_1^{'\ell}, \sqrt{g_{22}}x_2^\ell + z_2^{'\ell}) - h(z_1^{'\ell}, z_2^{'\ell}) + \ell\varepsilon_\ell \\
&\le \frac{\ell}{2}\log(1+g_{12}+g_{22}) + \ell\varepsilon_\ell, && \text{(A.17)}
\end{aligned}$$

where the data processing inequality implies (A.16) for the Markov chain $x_2^\ell \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow y_1^\ell$. Note that this bound is essentially the cut-set bound for the cut $\Omega_s = \{S_2\}$ and $\Omega_d = \{S_1, A, B, D_1, D_2\}$.

- (GZS-4) $R_2 < \frac{1}{2}\log(1+g_{12}) + \frac{1}{2}\log(1+h_{22})$: Again we use Lemma 6 to upper bound R_2 as

$$\begin{aligned}
\ell R_2 &\le I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell \\
&\le I(x_2^{'\ell}, x_2^\ell; x_1^\ell, y_1^{'\ell}, y_2^\ell) + \ell\varepsilon_\ell \\
&= I(x_2^{'\ell}, x_2^\ell; x_1^\ell) + I(x_2^{'\ell}, x_2^\ell; y_1^{'\ell}, y_2^\ell | x_1^\ell) + \ell\varepsilon_\ell \\
&= I(x_2^{'\ell}, x_2^\ell; y_1^{'\ell} | x_1^\ell) + I(x_2^{'\ell}, x_2^\ell; y_2^\ell | x_1^\ell, y_1^{'\ell}) + \ell\varepsilon_\ell \\
&= I(x_2^\ell; y_1^{'\ell} | x_1^\ell) + I(x_2^{'\ell}; y_1^{'\ell} | x_1^\ell, x_2^\ell) + I(x_2^{'\ell}; y_2^\ell | x_1^\ell, y_1^{'\ell}) + I(x_2^\ell; y_2^\ell | x_1^\ell, y_1^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell \quad \text{(A.18)} \\
&= I(x_2^\ell; y_1^{'\ell} | x_1^\ell) + I(x_2^{'\ell}; y_2^\ell | x_1^\ell, y_1^{'\ell}) + \ell\varepsilon_\ell \\
&\le \frac{\ell}{2}\log(1+g_{12}) + \frac{\ell}{2}\log(1+h_{22}) + \ell\varepsilon_\ell. \quad \text{(A.19)}
\end{aligned}$$

  Note that we used the fact that the second and fourth terms in (A.18) are zero. This follows from

$$I(x_2^{'\ell}; y_1^{'\ell} | x_1^\ell, x_2^\ell) \le I(x_2^{'\ell}; y_1^{'\ell} - \sqrt{g_{11}} x_1^\ell - \sqrt{g_{12}} x_2^\ell | x_1^\ell, x_2^\ell) = I(x_2^{'\ell}; z_1^{'\ell} | x_1^\ell, x_2^\ell) = 0,$$

  and

$$\begin{aligned}
I(x_2^\ell; y_2^\ell | x_1^\ell, y_1^{'\ell}, x_2^{'\ell}) &\le I(x_2^\ell; y_2^\ell | x_1^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&\le I(x_2^\ell; y_2^\ell - \sqrt{h_{21}} x_1^{'\ell} - \sqrt{h_{22}} x_2^{'\ell} | x_1^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&\le I(x_2^\ell; z_2^\ell | x_1^\ell, x_1^{'\ell}, x_2^{'\ell}) = 0.
\end{aligned}$$

- (GZS-5) $R_1 + R_2 < \frac{1}{2}\log(1+g_{22}) + \frac{1}{2}\log(1+h_{11}+h_{21})$:

  We start from Lemma 6 and write

$$\begin{aligned}
\ell(R_1+R_2) &\le I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \quad \text{(A.20)} \\
&\le I(y_1^\ell, y_2^\ell; x_1^{'\ell}, x_2^\ell) + \ell\varepsilon_\ell \\
&\le I(y_1^\ell, y_2^\ell, y_2^{'\ell}; x_1^{'\ell}, x_2^\ell) + \ell\varepsilon_\ell \\
&= I(y_2^{'\ell}; x_2^\ell) + I(y_2^{'\ell}; x_1^{'\ell} | x_2^\ell) + I(y_1^\ell, y_2^\ell; x_1^{'\ell}, x_2^\ell | y_2^{'\ell}) + \ell\varepsilon_\ell \\
&= I(y_2^{'\ell}; x_2^\ell) + I(y_1^\ell, y_2^\ell; x_1^{'\ell}, x_2^\ell | x_2^{'\ell}) + \ell\varepsilon_\ell \\
&= I(y_2^{'\ell}; x_2^\ell) + h(y_1^\ell, y_2^\ell | x_2^{'\ell}) - h(y_1^\ell, y_2^\ell | x_2^\ell, x_1^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell \\
&\le I(y_2^{'\ell}; x_2^\ell) + h(\sqrt{h_{11}} x_1^{'\ell} + z_1^\ell, \sqrt{h_{21}} x_1^{'\ell} + z_2^\ell) - h(z_1^\ell, z_2^\ell) + \ell\varepsilon_\ell \\
&\le \frac{\ell}{2}\log(1+g_{22}) + \frac{\ell}{2}\log(1+h_{11}+h_{21}) + \ell\varepsilon_\ell \quad \text{(A.21)}
\end{aligned}$$

where (A.20) follows from the data processing inequality for the Markov chain $(x_1^\ell, x_2^\ell) \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow (y_1^\ell, y_2^\ell)$. Note that this bound essentially captures the maximum flow of information through the cut $\Omega_s = \{S_1, S_2, A\}$ and $\Omega_d = \{B, D_1, D_2\}$.

- (GZS-6) $R_1 + R_2 < \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log(1 + h_{22})$:
Similar to the previous bounds, we start from Lemma 6 and write

$$\begin{aligned}
R_1 + R_2 &\le I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\
&\le I(y_1^{'\ell}, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell && \text{(A.22)} \\
&\le I(y_1^{'\ell}, y_2^\ell; x_1^\ell, x_2^\ell, x_2^{'\ell}) + \ell\varepsilon_\ell \\
&= I(y_1^{'\ell}; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_2^{'\ell}|x_1^\ell, x_2^\ell) + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + I(y_2^\ell; x_1^\ell, x_2^\ell|y_1^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell && \text{(A.23)} \\
&= I(y_1^{'\ell}; x_1^\ell, x_2^\ell) + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + \ell\varepsilon_\ell \\
&\le \frac{\ell}{2}\log(1 + g_{11} + g_{12}) + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + \ell\varepsilon_\ell. && \text{(A.24)}
\end{aligned}$$

Note that in (A.22) we used the data processing inequality. An argument similar to that is used in the proof of (GZS-4) shows that the second and fourth terms in (A.23) are zero. Now, we have

$$\begin{aligned}
I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) &= h(y_2^\ell|y_1^{'\ell}) - h(y_2^\ell|x_2^{'\ell}, y_1^{'\ell}) \\
&\le h(y_2^\ell|x_1^{'\ell}) - h(y_2^\ell|x_1^{'\ell}, x_2^{'\ell}, y_1^{'\ell}) \\
&= h(\sqrt{h_{22}}x_2^{'\ell} + z_2^\ell|x_1^{'\ell}) - h(z_2^\ell|x_1^{'\ell}, x_2^{'\ell}, y_1^{'\ell}) \\
&\le h(\sqrt{h_{22}}x_2^{'\ell} + z_2^\ell) - h(z_2^\ell) \\
&\le \frac{\ell}{2}\log(1 + h_{22}) && \text{(A.25)}
\end{aligned}$$

Finally, we obtain the desired bound by replacing (A.25) in (A.24). It is worth mentioning that this bound is the same as the cut-set bound for the cut $\Omega_s = \{S_1, S_2, B\}$ and $\Omega_d = \{A, D_1, D_2\}$.

- (GZS-7) $R_1 < \frac{1}{2}\log(1 + h_{11})$: Using Lemma 6 and the data processing inequality, we can write

$$\ell R_1 \le I(x_1^{'\ell}; y_1^\ell) + \ell\varepsilon_\ell \le I(x_1^{'\ell}; y_1^\ell) + \ell\varepsilon_\ell \le \frac{\ell}{2}\log(1 + h_{11}) + \ell\varepsilon_\ell. \quad \text{(A.26)}$$

- (GZS-8) $R_2 < \frac{1}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}})$: Starting from Lemma 6 and applying the data processing inequality for the Markov chain $x_2^\ell \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow y_2^\ell$, we have

$$\ell R_2 \le I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell \le I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) + \ell\varepsilon_\ell \le \frac{\ell}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}) + \ell\varepsilon_\ell. \quad \text{(A.27)}$$

Note that $x_1^{'\ell}$ and $x_2^{'\ell}$ are not independent. However, their variance is upper bounded by $(\sqrt{h_{21}} + \sqrt{h_{22}})^2$.

- (GZS-9) $R_2 < \frac{1}{2}\log(1+g_{22}) + \frac{1}{2}\log(1+h_{21})$: Consider the cut which partitions the network into $\Omega_s = \{S_1, S_2, A, D_1\}$ and $\Omega_d = \{B, D_2\}$. We have

$$\begin{aligned}
\ell R_2 &\leq I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell \\
&\leq I(x_1^{'\ell}, x_2^\ell; y_2^{'\ell}, y_2^\ell) + \ell\varepsilon_\ell \\
&= I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^{'\ell}|x_2^\ell) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}) + I(x_2^\ell; y_2^\ell|y_2^{'\ell}, x_1^{'\ell}) + \ell\varepsilon_\ell \qquad \text{(A.28)} \\
&= I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}) + \ell\varepsilon_\ell \\
&= I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+g_{22}) + \frac{\ell}{2}\log(1+h_{21}) + \ell\varepsilon_\ell. \qquad \text{(A.29)}
\end{aligned}$$

  We again used an argument similar to that is used in proof of (GZS-4) to show that the second and fourth terms in (A.28) are zero.

  This completes the proof of the outer bound in Theorem 2.

A.2 The Achievability Part

In this section we provide an encoding scheme for the Gaussian ZS network, and show that the rate region that can be achieved using this scheme is only a constant bit gap away from the outer bound.

Large Channel Gains: In this part, we assume that all channel gains are at least 1, *i.e.*, $g_{ij} \geq 1$, and $h_{ij} \geq 1$. Note that if any of the gains are small, then either one of the rates are small (of the order of our constant bit gap), or the cross links are negligible. We will discuss these cases later.

The encoding scheme proposed for the Gaussian ZS network consists of two separate parts. We first split the message of the second source nodes as $W_1 = U_1^{(1)}$ and $W_2 = (U_2^{(1)}, U_2^{(2)}, U_2^{(3)})$, where $U_2^{(1)}$ can be decoded at both relay nodes A and B, and $U_2^{(2)}$ and $U_2^{(3)}$ can be decoded only at A and B, respectively (see Figure 10). Denoting the rate of message $W_i^{(j)}$ by $\Upsilon_{i,j}$, the following rate constraints are imposed by this message splitting

$$R_1 = \Upsilon_{1,1}, \qquad \text{(A.30)}$$

$$R_2 = \Upsilon_{2,1} + \Upsilon_{2,2} + \Upsilon_{2,3}. \qquad \text{(A.31)}$$

An achievable rate region for this message splitting is given in Lemma 1.

In the second layer of the network (see Figure 11), relay node A further splits its messages as follows: $W_1 = U_1^{(1)} = \left(V_1^{(1)}, V_1^{(2)}\right)$, $U_2^{(1)} = \left(V_2^{(1)}, V_2^{(2)}\right)$, and $U_2^{(2)} = \left(V_2^{(3)}, V_2^{(4)}\right)$. A similar message splitting

is also performed at node B to obtain $U_2^{(1)} = \left(V_2^{(1)}, V_2^{(2)}\right)$ and $U_2^{(3)} = V_2^{(5)}$. This message splitting imposes the following rate equations

$$\Upsilon_{1,1} = \Theta_{1,1} + \Theta_{1,2}, \tag{A.32}$$

$$\Upsilon_{2,1} = \Theta_{2,1} + \Theta_{2,2}, \tag{A.33}$$

$$\Upsilon_{2,2} = \Theta_{2,3} + \Theta_{2,4}, \tag{A.34}$$

$$\Upsilon_{2,3} = \Theta_{2,5}, \tag{A.35}$$

where $\Theta_{i,j}$ denotes the rate of the message $V_i^{(j)}$. Next, the relay nodes have to convey the messages to the destination nodes such that D_1 can decode $V_1^{(1)}$, $V_1^{(2)}$, $V_2^{(1)}$ and $V_2^{(3)}$, and D_2 be able to decode $V_1^{(1)}$, $V_2^{(1)}$, $V_2^{(2)}$, $V_2^{(3)}$, $V_2^{(4)}$ and $V_2^{(5)}$. An achievable rate region for this transmission scenario is given in Lemma 2.

Putting the rate constraints in Lemma 1 and Lemma 2 together with the equations in (A.30)-(A.31)

and (A.32)-(A.35), we obtain the following achievable rate region for the Gaussian ZS network.

$$\mathcal{R}_{\text{ach}}^{\mathsf{GZS}} = \Big\{(R_1,R_2) : \exists \Upsilon_{1,1},\Upsilon_{2,1},\Upsilon_{2,2},\Upsilon_{2,3},\Theta_{1,1},\Theta_{1,2},\Theta_{2,1},\Theta_{2,2},\Theta_{2,3},\Theta_{2,4},\Theta_{2,5} \geq 0, \tag{A.36}$$

$$R_1 = \Upsilon_{1,1},$$

$$R_2 = \Upsilon_{2,1} + \Upsilon_{2,2} + \Upsilon_{2,3},$$

$$\Upsilon_{1,1} = \Theta_{1,1} + \Theta_{1,2},$$

$$\Upsilon_{2,1} = \Theta_{2,1} + \Theta_{2,2},$$

$$\Upsilon_{2,2} = \Theta_{2,3} + \Theta_{2,4},$$

$$\Upsilon_{2,3} = \Theta_{2,5},$$

$$\Upsilon_{1,1} \leq \left(\frac{1}{2}\log\left(1+g_{11}\right) - \frac{1}{2}\right)^+,$$

$$\Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+\frac{g_{12}}{g_{22}}\right) - \frac{1}{2}\right)^+,$$

$$\Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+g_{12}\right) - \frac{1}{2}\right)^+,$$

$$\Upsilon_{1,1} + \Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+g_{11}+g_{12}\right) - \frac{1}{2}\right)^+,$$

$$\Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\right)^+,$$

$$\Upsilon_{2,1} + \Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1+g_{22}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{1,1} + \Theta_{1,2} + \Theta_{2,1} + \Theta_{2,3} \leq \left(\frac{1}{2}\log\left(1+h_{11}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{1,2} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{11}}{h_{12}}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{21}}{h_{11}}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{1,1} + \Theta_{2,3} + \Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+h_{21}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1+h_{22}\right) - \frac{1}{2}\right)^+,$$

$$\Theta_{1,1} + \Theta_{2,1} + \Theta_{2,2} + \Theta_{2,3} + \Theta_{2,4} + \Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1+h_{21}+h_{22}\right) - \frac{1}{2}\right)^+ \Big\}.$$

We apply the Fourier-Motzkin elimination on this region, to project it on the coordinated R_1 and R_2, and obtain the following rate region. After some simplifications, we get

$$\mathcal{R}_{\text{ach}}^{\mathsf{GZS}} = \Big\{(R_1,R_2): R_1 \le \left(\frac{1}{2}\log\left(g_{11}\right)-\frac{1}{2}\right)^+,$$
$$R_2 \le \left(\frac{1}{2}\log\left(g_{12}+g_{22}\right)-\frac{1}{2}\right)^+,$$
$$R_1+R_2 \le \left(\frac{1}{2}\log\left(g_{11}+g_{12}\right)+\frac{1}{2}\log\left(\frac{g_{22}}{g_{12}}\right)-\frac{1}{2}\right)^+,$$
$$R_2 \le \left(\frac{1}{2}\log\left(g_{12}\right)+\frac{1}{2}\log\left(h_{22}\right)-\frac{1}{2}\right)^+,$$
$$R_1+R_2 \le \left(\frac{1}{2}\log\left(g_{22}\right)+\frac{1}{2}\log\left(h_{11}+h_{21}\right)-\frac{1}{2}\right)^+,$$
$$R_1+R_2 \le \left(\frac{1}{2}\log\left(g_{11}+g_{12}\right)+\frac{1}{2}\log\left(h_{22}\right)-\frac{1}{2}\right)^+,$$
$$R_1 \le \left(\frac{1}{2}\log\left(h_{11}\right)-\frac{1}{2}\right)^+,$$
$$R_2 \le \left(\frac{1}{2}\log\left(h_{21}+h_{22}\right)-\frac{1}{2}\right)^+,$$
$$R_2 \le \left(\frac{1}{2}\log\left(g_{22}\right)+\frac{1}{2}\log\left(h_{21}\right)-\frac{1}{2}\right)^+,$$
$$R_1+R_2 \le \left(\frac{1}{2}\log\left(h_{21}+h_{22}\right)+\frac{1}{2}\log\left(\frac{h_{11}}{h_{21}}\right)-\frac{1}{2}\right)^+\Big\}.$$

Note that this rate region is characterized by a set of constraints which are similar to the inequalities in the definition of $\mathcal{R}^{\mathsf{GZS}}$, except for the additive constants, and the fact that $\log(1+x)$ is replaced by $\log(x)$. Note that since $x \ge 1$, we have

$$\frac{1}{2}\log(1+x)-\frac{1}{2}\log(x) \le \frac{1}{2}. \tag{A.37}$$

Hence, the difference between the RHS's of two sets of inequalities do not exceed 1 for R_1, and $3/2$ for R_2 and R_1+R_2. Therefore, for any rate pair $(R_1,R_2) \in \mathcal{R}^{\mathsf{GZS}}$, we have $(R_1-1,R_2-1.5) \in \mathcal{R}_{\text{ach}}^{\mathsf{GZS}}$. This completes the proof.

Small Channel Gains: We will show in this part that if any of the channel gains are small, then the outer bound in Theorem 2 is still within a constant bit gap of an achievable rate region. This argument is based on the analysis of the same network, in which all the links with gain smaller than 1 are removed. One can show that the capacity region of this modified network is within a constant gap from that of the original one. On the other hand, we can argue that the gap between the achievable rate pairs of the modified network and the outer bound in Theorem 2 is bounded by a constant. Therefore, we can

conclude that if $(R_1, R_2) \in \mathcal{R}^{\mathsf{GZS}}$ then $(R_1 - \delta_1, R_2 - \delta_2)$ is achievable for the original network, where $\delta_1 = 1$ and $\delta_2 = 1.5$.

The main intuition behind this argument is the fact that since all the nodes are assumed to have power constraint equal to 1, the flow of information through a link with gain not exceeding 1 is upper bounded by $\frac{1}{2}\log(1+\mathsf{SNR}) \leq \frac{1}{2}\log(1+1) = \frac{1}{2}$ bit. Therefore, by removing such links from the network, the achievable rates change by at most $\frac{1}{2}$ bit. On the other hand, the incoming signals over small channel gains may act as an interference on the original network, which cause a total noise power not exceeding 1. Therefore, by doubling the noise variances of the original network, we guarantee that capacity region of the modified network is always smaller than that of the original one.

The advantage of analyzing the modified network instead of the original one is that some of the links are removed in the modified network, which convert it to simpler network to analyze.

A precise analysis of the modified networks requires considering several cases separately. However, similar techniques and ideas will be used for all cases. In the following we present one illustrating example, and skip the details for the other cases.

Example 5: Consider the Gaussian ZS network in Figure 3(a), and assume that $g_{12} = 0$. Therefore, the first layer of the network would be two parallel links as shown in Figure 16, where $\mathbb{E}[\tilde{z}_1'^2] = 2$. Moreover, the rate region in (GZS-1)-(GZS-10) will be reduced to

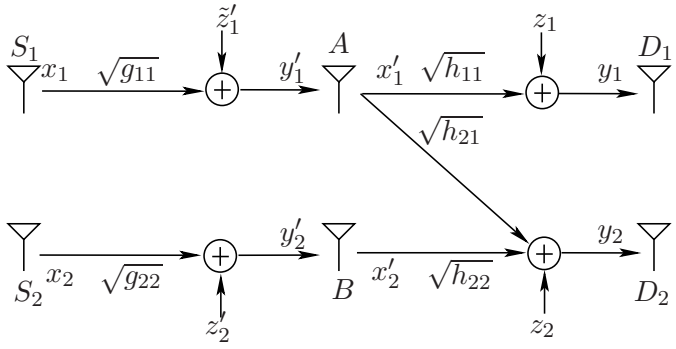


Fig. 16. A modified ZS network obtained assuming $g_{12} = 0$.

$$R_1 \leq \frac{1}{2}\log(1+g_{11}) \tag{A.38}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{12}) \tag{A.39}$$

$$R_1 \leq \frac{1}{2}\log(1+h_{11}) \tag{A.40}$$

$$R_2 \leq \frac{1}{2}\log(1+h_{22}) \tag{A.41}$$

$$R_1+R_2 \leq \frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}})+\frac{1}{2}\log\left(1+\frac{h_{11}}{h_{21}}\right). \tag{A.42}$$

The encoding strategy for this network is fairly simple. Let (R_1, R_2) be a rate pair satisfying (A.38)-(A.42). The goal is to show that (R_1-1, R_2-1) is achievable. Since (R_1-1, R_2-1) satisfies (A.38) and (A.39), transmission over the first layer of the network from the source nodes to the relays is simply done using random Gaussian codes.

The second layer of the network is a Gaussian S network. Once the relays decode the messages received from the first layer of the network, they encode them using an encoding strategy similar to that of the Z network in Example 4 in Section III. Note that the sum-rate bounds in (A.42) and the outer bound of the S network are slightly different. However, their difference is upper bounded by

$$\begin{aligned}\frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}})-\frac{1}{2}\log(1+h_{21}+h_{22}) &= \frac{1}{2}\log\left(1+\frac{2\sqrt{h_{21}h_{22}}}{1+h_{21}+h_{22}}\right)\\ &< \frac{1}{2}\log(1+1)=\frac{1}{2}.\end{aligned} \tag{A.43}$$

Therefore, the loss caused by this difference is at most $\frac{1}{2}$ bit, and (R_1-1, R_2-1) would be achievable. On the other hand, as we argued before, the capacity of the modified network is an inner bound for the original one, and hence, (R_1-1, R_2-1) is achievable for the ZS network as well.

APPENDIX B
THE GAUSSIAN ZZ NETWORK

B.1 The Outer Bound

In the following we present the proof for each of the inequalities in (GZZ1)-(GZZ6), separately. We again present the Gaussian ZZ network in Figure 17, to clarify the notation used in the proof. In particular, we use two variables, which are the noisy signals received at A and D_1 through the cross links assuming the direct links were absent, namely,

$$\gamma_1 = \sqrt{g_{12}}x_2 + z_1',$$

$$\gamma_2 = \sqrt{h_{12}}x_2' + z_1.$$

Note that $y_1' = \sqrt{g_{11}}x_1 + \gamma_1$ and $y_1 = \sqrt{h_{11}}x_1' + \gamma_2$.

Suppose that the rate pair (R_1, R_2) is achieved with a small decoding error probability ε_ℓ using a code of length ℓ. The following chains of inequalities provide upper bounds on the individual rates as well as the sum-rate. We again use Lemma 6, which essentially captures the decodability requirements of the network.

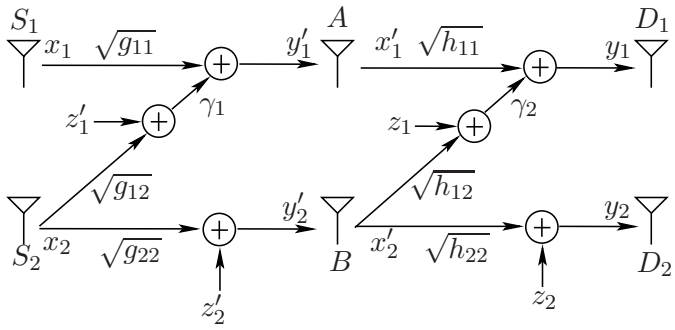


Fig. 17. The Gaussian ZZ network.

The individual rate bounds in (GZZ1)-(GZZ4) have the same structure as the cut-set bound, although we derive them through a slightly different argument. However, the two sum-rate bounds in (GZZ5) and (GZZ6) are conceptually different than the cut-set bounds. These two bounds which are tighter than cut-set bounds are derived through a genie-aided argument; that is, we assume that the signal sent over the cross link of one layer is given by a genie to the receiver of the other layer (relay node A in layer 1 and destination node D_1 in layer 2). Therefore, we present the proofs of (GZZ5) and (GZZ6) first. The more standard cut-set type bounds are provided later for completeness.

a) The proof of the genie-aided bounds:

- (GZZ5) $R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) + \frac{1}{2}\log(1 + h_{12})$: We start with the

sum-rate inequality in Lemma 6, and write

$$\begin{aligned}
\ell(R_1+R_2) &\le I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\
&\le I(y_1^{'\ell}, y_2^{'\ell}; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\
&\le I(y_1^{'\ell}, y_2^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_2^{'\ell}; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + \ell\varepsilon_\ell \\
&\le I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_2^{'\ell}, \gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + \ell\varepsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + I(y_2^{'\ell}; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) + \ell\varepsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + I(y_2^{'\ell}; x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) + I(y_2^{'\ell}; x_1^\ell | x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) + \ell\varepsilon_\ell.
\end{aligned} \tag{B.1}$$

Each of the terms in (B.1) can be bounded as follows. In order to bound the first term, we can simply write

$$\begin{aligned}
I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) &= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell | \gamma_2^\ell) \\
&= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + h(y_1^{'\ell}|\gamma_2^\ell) - h(y_1^{'\ell}|x_1^\ell, x_2^\ell, \gamma_2^\ell) \\
&\le I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + h(y_1^{'\ell}) - h(y_1^{'\ell}|x_1^\ell, x_2^\ell) && \text{(B.2)} \\
&= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell) \\
&\le I(\gamma_2^\ell; x_2^{'\ell}) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell) && \text{(B.3)} \\
&= \frac{\ell}{2}\log(1+h_{12}) + \frac{\ell}{2}\log(1+g_{11}+g_{12}), && \text{(B.4)}
\end{aligned}$$

where in (B.2) we have used the fact that conditioning decreases the entropy, and the Markov chain $\gamma_2^\ell \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^{'\ell} \leftrightarrow (x_1^\ell, x_2^\ell) \leftrightarrow y_1^{'\ell}$. Also (B.3) follows from the same Markov chain.

For the second term, we can write

$$\begin{aligned}
I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) &= I(y_1^{'\ell} - \gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&= I(\sqrt{g_{11}}x_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&\le I(W_1; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&\le H(W_1 | y_1^{'\ell}, \gamma_2^\ell) \\
&\le H(W_1 | y_1^\ell) \le \ell\varepsilon_\ell, && \text{(B.5)}
\end{aligned}$$

where the last inequality holds since $y_1^\ell = \sqrt{h_{11}}x_1^{'\ell} + \gamma_2^\ell = f_1(y_1^{'\ell}) + \gamma_2^\ell = f_2(y_1^{'\ell}, \gamma_2^\ell)$.

In order to bound the third term in (B.1) we can write

$$
\begin{aligned}
I(y_2^{'\ell}; x_2^{\ell}|y_1^{'\ell}, \gamma_1^{\ell}, \gamma_2^{\ell}) &= h(x_2^{\ell}|y_1^{'\ell}, \gamma_1^{\ell}, \gamma_2^{\ell}) - h(x_2^{\ell}|y_1^{'\ell}, y_2^{'\ell}, \gamma_1^{\ell}, \gamma_2^{\ell}) \\
&\leq h(x_2^{\ell}|\gamma_1^{\ell}) - h(x_2^{\ell}|y_1^{'\ell}, y_2^{'\ell}, \gamma_1^{\ell}, \gamma_2^{\ell}) && \text{(B.6)} \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - [h(x_2^{\ell}, t_1^{\ell}|y_1^{'\ell}, y_2^{'\ell}, \gamma_2^{\ell}) - h(t_1^{\ell}|y_1^{'\ell}, y_2^{'\ell}, \gamma_2^{\ell})] \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - [h(x_2^{\ell}, t_1^{\ell}|y_1^{'\ell}, y_2^{'\ell}) - h(t_1^{\ell}|y_1^{'\ell}, y_2^{'\ell})] && \text{(B.7)} \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - h(x_2^{\ell}|y_1^{'\ell}, y_2^{'\ell}, \gamma_1^{\ell}) \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - h(x_2^{\ell}|x_1^{\ell}, y_2^{'\ell}, \gamma_1^{\ell}) && \text{(B.8)} \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - [h(x_2^{\ell}|y_2^{'\ell}, \gamma_1^{\ell}) + h(x_1^{\ell}|x_2^{\ell}, y_2^{'\ell}, \gamma_1^{\ell}) - h(x_1^{\ell}|y_2^{'\ell}, \gamma_1^{\ell})] \\
&= h(x_2^{\ell}|\gamma_1^{\ell}) - h(x_2^{\ell}|y_2^{'\ell}, \gamma_1^{\ell}) && \text{(B.9)} \\
&= I(y_2^{'\ell}; x_2^{\ell}|\gamma_1^{\ell}) \\
&= h(y_2^{'\ell}|\gamma_1^{\ell}) - h(y_2^{'\ell}|\gamma_1^{\ell}, x_2^{\ell}) \\
&= h(y_2^{'\ell}|\gamma_1^{\ell}) - h(y_2^{'\ell}|x_2^{\ell}) && \text{(B.10)} \\
&= h(y_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}\gamma_1^{\ell}|\gamma_1^{\ell}) - h(y_2^{'\ell} - \sqrt{g_{22}}x_2^{\ell}|x_2^{\ell}) \\
&= h(z_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}z_1^{'\ell}|\gamma_1^{\ell}) - h(z_2^{'\ell}|x_2^{\ell}) \\
&\leq h(z_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}z_1^{'\ell}) - h(z_2^{'\ell}) && \text{(B.11)} \\
&= \frac{\ell}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right), && \text{(B.12)}
\end{aligned}
$$

where in (B.6) we have used the fact that conditioning reduces the differential entropy, and (B.7) holds due to the Markov chain $(x_2^{\ell}, t_1^{\ell}) \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow \gamma_2^{\ell}$. Then in (B.8) we replaced $(y_1^{'\ell}, \gamma_1^{\ell})$ by $(x_1^{\ell}, \gamma_1^{\ell})$ since there is an one-to-one map, $y_1^{'\ell} = \sqrt{g_{11}}x_1^{\ell} + t_1^{\ell}$, between these joint variables, and in (B.9) we used the fact that x_1^{ℓ} is independent of $(x_2^{\ell}, y_2^{'\ell}, t_1^{\ell})$ to conclude $h(x_1^{\ell}|x_2^{\ell}, y_2^{'\ell}, \gamma_1^{\ell}) = h(x_1^{\ell}|y_2^{'\ell}, \gamma_1^{\ell}) = h(x)$. Also (B.10) holds due to the Markov chain $y_2^{'\ell} \leftrightarrow x_2^{\ell} \leftrightarrow \gamma_1^{\ell}$. Finally, (B.11) is true due to removing conditioning and the fact that z_2' is independent of x_2.

Finally for the last term in (B.1) we have

$$\begin{aligned} I(y_2^{'\ell}; x_1^\ell | x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) &\leq I(y_2^{'\ell}; W_1 | x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\ &\leq H(W_1 | x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\ &\leq H(W_1 | y_1^{'\ell}, \gamma_2^\ell) \\ &\leq H(W_1 | y_1^\ell) \qquad \text{(B.13)} \\ &\leq \ell\varepsilon_\ell, \qquad \text{(B.14)} \end{aligned}$$

where (B.13) is due to the fact that $y_1^\ell = \sqrt{h_{11}} x_1^{'\ell} + \gamma_2^\ell = f_1(y_1^{'\ell}) + \gamma_2^\ell = f_2(y_1^{'\ell}, \gamma_2^\ell)$ is a function of $(y_1^{'\ell}, \gamma_2^\ell)$, and (B.14) is just the Fano's inequality.

Replacing (B.4), (B.5), (B.12), and (B.14) in (B.1), we get

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) + \frac{1}{2}\log(1 + h_{12}) + 3\ell\varepsilon_\ell. \qquad \text{(B.15)}$$

- (GZZ6) Before proving this inequality, we present a lemma which will be used in this proof. We will present the proof of this lemma later in Appendix C.

  ***Lemma** 7:* Let X_1 and X_2 be two (arbitrarily correlated) random variables with variance constraints $\mathbb{E}[X_1^2] = \sigma_1^2$ and $\mathbb{E}[X_2^2] = \sigma_2^2$, which form a Markov chain $X_1 \leftrightarrow \Gamma \leftrightarrow X_2$ for some random variable Γ. Also assume that Z is a zero-mean unit variance Gaussian random variable independent of X_1, X_2 and Γ. Then the conditional differential entropy of $Y = X_1 + X_2 + Z$ is upper bounded by

$$h(Y|\Gamma) \leq \frac{1}{2}\log 2\pi e(1 + \sigma_1^2 + \sigma_2^2). \qquad \text{(B.16)}$$

Now, in order to prove (GZZ5), we start with Lemma 6.

$$\begin{aligned} \ell(R_1 + R_2) &\leq I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\ &\leq I(y_1^\ell, y_2^\ell, \gamma_1^\ell; x_1^\ell, x_2^\ell) + \ell\varepsilon_\ell \\ &= I(y_1^\ell, \gamma_1^\ell; x_1^\ell, x_2^\ell) + I(y_2^\ell; x_1^\ell, x_2^\ell | y_1^\ell, \gamma_1^\ell) + \ell\varepsilon_\ell \\ &= I(\gamma_1^\ell; x_1^\ell, x_2^\ell) + I(y_1^\ell; x_1^\ell, x_2^\ell | \gamma_1^\ell) + I(y_2^\ell; x_1^\ell, x_2^\ell | y_1^\ell, \gamma_1^\ell) + \ell\varepsilon_\ell. \qquad \text{(B.17)} \end{aligned}$$

Since γ_1^ℓ is independent of x_1^ℓ, the first term can be simply bounded as

$$I(\gamma_1^\ell; x_1^\ell, x_2^\ell) = I(\gamma_1^\ell; x_2^\ell) + I(t^\ell; x_1^\ell | x_2^\ell) = I(\gamma_1^\ell; x_2^\ell) + I(z_1^{'\ell}; x_1^\ell | x_2^\ell) \leq \frac{\ell}{2}\log(1 + g_{12}). \qquad \text{(B.18)}$$

For the second term we can write

$$\begin{aligned} I(y_1^\ell; x_1^\ell, x_2^\ell|\gamma_1^\ell) &= h(y_1^\ell|\gamma_1^\ell) - h(y_1^\ell|x_1^\ell, x_2^\ell, \gamma_1^\ell) \\ &\leq h(y_1^\ell|\gamma_1^\ell) - h(y_1^\ell|x_1^{'\ell}, x_2^{'\ell}) \qquad \text{(B.19)} \\ &= h(y_1^\ell|\gamma_1^\ell) - h(z_1^\ell|x_1^{'\ell}, x_2^{'\ell}) \\ &= h(\sqrt{h_{11}}x_1^{'\ell} + \sqrt{h_{12}}x_2^{'\ell} + z_1^\ell|\gamma_1^\ell) - h(z_1^\ell) \\ &\leq \frac{\ell}{2}\log\Big(2\pi e(1 + h_{11} + h_{12})\Big) - \frac{\ell}{2}\log 2\pi e \qquad \text{(B.20)} \\ &= \frac{\ell}{2}\log(1 + h_{11} + h_{12}), \qquad \text{(B.21)} \end{aligned}$$

where (B.19) follows from the Markov chain $y_1^\ell \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow (x_1^\ell, x_2^\ell, \gamma_1^\ell)$. In (B.20) we have used Lemma 7 for t_1^ℓ, $x_1^{'\ell}$ and $x_2^{'\ell}$ which form a Markov chain, since

$$I(x_1^{'\ell}; x_2^{'\ell}|t_1^\ell) \leq I(y_1^{'\ell}; y_2^{'\ell}|t_1^\ell) = I(\frac{y_1^{'\ell} - t_1^\ell}{\sqrt{g_{11}}}; y_2^{'\ell}|t_1^\ell) = I(x_1^\ell; y_2^{'\ell}|t_1^\ell) = h(x_1^\ell|t_1^\ell) - h(x_1^\ell|t_1^\ell, y_2^{'\ell}) = 0.$$

The third term can be further upper bounded by

$$\begin{aligned} I(y_2^\ell; x_1^\ell, x_2^\ell|y_1^\ell, \gamma_1^\ell) &= h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_1^\ell, x_2^\ell, y_1^\ell, \gamma_1^\ell) \\ &\leq h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_2^{'\ell}) \qquad \text{(B.22)} \\ &\leq h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_2^{'\ell}, x_1^{'\ell}, y_1^\ell, \gamma_1^\ell) \qquad \text{(B.23)} \\ &= I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\ &\leq I(y_2^\ell, \gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\ &= I(\gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) + I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell), \qquad \text{(B.24)} \end{aligned}$$

where both (B.22) and (B.23) follow from the Markov chain $y_2^\ell \leftrightarrow x_2^{'\ell} \leftrightarrow (x_1^\ell, y_1^\ell, \gamma_1^\ell)$. Now, we have

$$
\begin{aligned}
I(\gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) &= I(y_1^\ell - \gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&= I(\sqrt{h_{11}}x_1^{'\ell}; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le I(y_1^{'\ell}; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \qquad \text{(B.25)} \\
&\le I(y_1^{'\ell} - t_1^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&= I(\sqrt{g_{11}}x_1^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le I(W_1; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le H(W_1|y_1^\ell) \le \ell\varepsilon_\ell, \qquad \text{(B.26)}
\end{aligned}
$$

where (B.25) follows from the fact that $x_1^{'\ell}$ is a function of $y_1^{'\ell}$. Finally,

$$
\begin{aligned}
I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) &= h(y_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) - h(y_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&= h(y_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}\gamma_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) - h(y_2^\ell - \sqrt{h_{22}}x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&= h(z_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}z_1^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) - h(z_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&\le h(z_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}z_1^\ell) - h(z_2^\ell) \qquad \text{(B.27)} \\
&= \frac{\ell}{2}\log\left(1 + \frac{h_{22}}{h_{12}}\right). \qquad \text{(B.28)}
\end{aligned}
$$

Here, in (B.27) we have used the fact that conditioning decreases the differential entropy, and the fact that z_2^ℓ is independent of $(y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell})$. Replacing (B.18), (B.21), (B.26), and (B.28) in (B.17), we will obtain the desired inequality.

b) The proofs of cut-set type bounds:

- (GZZ1) $R_1 \leq \frac{1}{2}\log(1+g_{11})$: The individual rate bound can be simply obtained from

$$\begin{aligned}
\ell R_1 &= I(x_1^\ell; y_1^\ell) + \ell\varepsilon_\ell \\
&\leq I(x_1^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\varepsilon_\ell \qquad \text{(B.29)} \\
&= I(x_1^\ell; y_1^{'\ell}|y_2^{'\ell}) + I(x_1^\ell; y_2^{'\ell}) + \ell\varepsilon_\ell \\
&= h(y_1^{'\ell}|y_2^{'\ell}) - h(y_1^{'\ell}|x_1^\ell, y_2^{'\ell}) \qquad \text{(B.30)} \\
&\leq h(y_1^{'\ell}|x_2^\ell) - h(z_1^{'\ell}|x_1^\ell, y_2^{'\ell}) \\
&= h(\sqrt{g_{11}}x_1^\ell + z_1^{'\ell}) - h(z_1^{'\ell}) + \ell\varepsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+g_{11}) + \ell\varepsilon_\ell,
\end{aligned}$$

  where (B.29) follows from the data processing inequality for the Markov chain $x_1^\ell \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow y_1^\ell$, and (B.30) follows from the Markov chain $y_1^{'\ell} \leftrightarrow x_2^\ell \leftrightarrow y_2^{'\ell}$). Note that $\varepsilon_\ell \to 0$ as ℓ grows. It is worth mentioning that this bound is similar to the cut-set bound for the cut $\Omega_s = \{S_1\}$ and $\Omega_d = \{S_2, A, B, D_1, D_2\}$.

- (GZZ2) $R_2 \leq \frac{1}{2}\log(1+g_{22})$:

  For the second rate bound, we can start with Lemma 6 and write

$$\ell R_2 \leq I(x_2^\ell; y_2^\ell) + \ell\varepsilon_\ell \leq I(x_2^\ell; y_2^{'\ell}) + \ell\varepsilon_\ell \leq \frac{\ell}{2}\log(1+g_{22}) + \ell\varepsilon_\ell,$$

  where we have used the data processing inequality and the Markov chain $x_2^\ell \leftrightarrow y_2^{'\ell} \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^\ell$ in the second inequality. Note that this bound captures the maximum flow of information through the cut specified by $\Omega_s = \{S_2\}$ and $\Omega_d = \{S_1, A, B, D_1, D_2\}$.

- (GZZ3) $R_1 \leq \frac{1}{2}\log(1+h_{11})$: In order to prove this upper bound, we use the cut-set bound for the cut $\Omega_s = \{S_1, S_2, A, B, D_2\}$ and $\Omega_d = \{D_1\}$.

$$\begin{aligned}
\ell R_1 &\leq I(x_1^{'\ell}; y_1^\ell|x_2^{'\ell}) + \ell\varepsilon_\ell \\
&= h(y_1^\ell|x_2^{'\ell}) - h(y_1^\ell|x_1^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell \\
&= h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell|x_2^{'\ell}) - h(z_1^\ell|x_1^{'\ell}, x_2^{'\ell}) + \ell\varepsilon_\ell \\
&\leq h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell) - h(z_1^\ell) + \ell\varepsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+h_{11}) + \ell\varepsilon_\ell.
\end{aligned}$$

- (GZZ4) $R_2 \leq \frac{1}{2}\log(1+h_{22})$: Starting from Lemma 6, we can write

$$\ell R_2 \leq I(x_2^{'\ell}; y_2^\ell) + \ell\varepsilon_\ell \leq I(x_2^{'\ell}; y_2^\ell) + \ell\varepsilon_\ell \leq \frac{1}{2}\log(1+h_{22}) + \ell\varepsilon_\ell,$$

where the second inequality follows from the data processing inequality for the Markov chain $x_2^\ell \leftrightarrow y_2^{'\ell} \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^\ell$.

This shows that the rate region in Theorem 4 is an outer bound for the achievable region region of the Gaussian ZZ network.

B.2 The Achievability Part

In this section we present an encoding/decoding scheme, and derive an achieve rate region for this strategy. We then show that the gap between the boundary of this achievable rate region and that of the outer bound presented in Theorem 4 is upper bounded by a constant.

Similar to the Gaussian ZS network, we only consider the large channel gain case, where we assume that all the channel gains are lower bounded by 1. A similar argument to that we used for the ZS network shows that for small channel gain cases the network is reduced to a simple one and its gap analysis is fairly simple.

We essentially use the result of Lemma 3 as an achievable rate region for the Z-neutralization network. We use notation (λ_g, μ_g) and (λ_h, μ_h) to distinguish between λ and μ parameters of the first and the second layers of the network.

In the first layer of the network, each source node splits its message into two parts, namely, functional and private parts, $W_1 = \left(U_1^{(0)}, U_1^{(1)}\right)$ and $W_2 = \left(U_2^{(0)}, U_2^{(1)}\right)$, where the functional parts, have the same rate, *i.e.,* $\Upsilon_{1,0} = \Upsilon_{2,0} = \Upsilon_0$. Both transmitters use a common lattice code to encode their functional sub-messages into $\mathbf{x}_{1,0} = \psi(U_1^{(0)})$ and $\mathbf{x}_{2,0} = \psi(U_2^{(0)})$, where ψ is the one-to-one encoding map induced by the lattice code. We define the partial-invertible function by

$$\phi\left(U_1^{(0)}, U_2^{(0)}\right) = \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)})\right) = \psi^{-1}(\mathbf{x}_{1,0} + \mathbf{x}_{2,0}). \tag{B.31}$$

We denote the rates of the private sub-messages by Υ_1 and Υ_2, where $\Upsilon_i = R_i - \Upsilon_0$, for $i = 1, 2$. The goal is to encode and forward messages to A and B in such a way that A can decode $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$, and B can decode $U_2^{(0)}$ and $U_2^{(1)}$. Based on Lemma 3, this can be done provided that

$$\begin{aligned}
\Upsilon_0 &\le \left(\frac{1}{2}\log(\lambda_g) - \frac{1}{2}\right)^+, \\
\Upsilon_0 + \Upsilon_1 &\le \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \\
\Upsilon_0 + \Upsilon_2 &\le \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \\
\Upsilon_0 + \Upsilon_1 + \Upsilon_2 &\le \left(\frac{1}{2}\log(\mu_g) - \frac{3}{2}\right)^+.
\end{aligned}$$

The second layer of the network is another Z-neutralization network with transmitters A and B, and receivers D_1 and D_2. We use $V_1^{(0)} = \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)})\right)$, $V_1^{(1)} = U_1^{(1)}$ as the functional and private messages of the first relay node, and $V_2^{(0)} = \psi^{-1}(-\mathbf{x}_{2,0}) = \psi^{-1}(-\psi(U_2^{(0)}))$ and $V_2^{(1)} = U_2^{(1)}$ for the functional and private messages of second relay. Denoting the corresponding rates by Θ_0, Θ_1, and Θ_2, we have

$$\Theta_i = \Upsilon_i, \qquad i = 0, 1, 2. \tag{B.32}$$

The goal is to encode and send these messages to the destinations, such that D_1 can decode $\phi(V_1^{(0)}, V_2^{(0)})$ and $V_1^{(1)}$, and D_2 can decode $V_2^{(0)}$ and $V_2^{(1)}$. Again we use the achievable rate region proposed in Lemma 3.

$$\begin{aligned}
\Theta_0 &\leq \left(\frac{1}{2}\log(\lambda_h) - \frac{1}{2}\right)^+, \\
\Theta_0 + \Theta_1 &\leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+, \\
\Theta_0 + \Theta_2 &\leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+, \\
\Theta_0 + \Theta_1 + \Theta_2 &\leq \left(\frac{1}{2}\log(\mu_h) - \frac{3}{2}\right)^+.
\end{aligned}$$

Note that the first destination observes $\phi(V_1^{(0)}, V_2^{(0)})$, which is equivalent to

$$\begin{aligned}
\phi\left(V_1^{(0)}, V_2^{(0)}\right) &= \psi^{-1}\left(\psi(V_1^{(0)}) + \psi(V_2^{(0)})\right) \\
&= \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)}) - \psi(U_2^{(0)})\right) \qquad \text{(B.33)} \\
&= U_1^{(0)}. \qquad \text{(B.34)}
\end{aligned}$$

Therefore, combining it with $V_1^{(1)} = U_1^{(1)}$, the first destination node can decode W_1. The second destination node D_2 has $V_2^{(0)}$ and $V_2^{(1)} = U_2^{(1)}$, and can compute

$$\psi^{-1}\left(-\psi(V_2^{(0)})\right) = U_2^{(0)}, \tag{B.35}$$

and hence it decodes W_2.

This scheme can reliably transmit the messages with rate pair in

$$\begin{aligned}\mathcal{R}_{\text{ach}}^{\mathsf{GZZ}} = \Big\{(R_1,R_2) :\exists\ \Upsilon_0,\Upsilon_1,\Upsilon_2,\Theta_0,\Theta_1,\Theta_2 \geq 0, \qquad &\text{(B.36)}\\ R_1 = \Theta_0+\Theta_1,\\ R_2 = \Theta_0+\Theta_2,\\ \Theta_i = \Upsilon_i, \qquad i=0,1,2,\\ \Upsilon_0 \leq \left(\frac{1}{2}\log\left(\lambda_g\right)-\frac{1}{2}\right)^+,\\ \Upsilon_0+\Upsilon_1 \leq \left(\frac{1}{2}\log\left(g_{11}\right)-1\right)^+,\\ \Upsilon_0+\Upsilon_2 \leq \left(\frac{1}{2}\log\left(g_{22}\right)-1\right)^+,\\ \Upsilon_0+\Upsilon_1+\Upsilon_2 \leq \left(\frac{1}{2}\log\left(\mu_g\right)-\frac{3}{2}\right)^+,\\ \Theta_0 \leq \left(\frac{1}{2}\log\left(\lambda_h\right)-\frac{1}{2}\right)^+,\\ \Theta_0+\Theta_1 \leq \left(\frac{1}{2}\log\left(h_{11}\right)-1\right)^+,\\ \Theta_0+\Theta_2 \leq \left(\frac{1}{2}\log\left(h_{22}\right)-1\right)^+,\\ \Theta_0+\Theta_1+\Theta_2 \leq \left(\frac{1}{2}\log\left(\mu_h\right)-\frac{3}{2}\right)^+\Big\}. \qquad &\text{(B.37)}\end{aligned}$$

It only remains to apply Fourier-Motzkin elimination to project this region onto (R_1,R_2). This gives us

$$\begin{aligned}\mathcal{R}_{\text{ach}}^{\mathsf{GZZ}} = \Big\{(R_1,R_2) :&R_1 \leq \left(\frac{1}{2}\log\left(g_{11}\right)-1\right)^+,\\ &R_1 \leq \left(\frac{1}{2}\log\left(h_{11}\right)-1\right)^+,\\ &R_2 \leq \left(\frac{1}{2}\log\left(g_{22}\right)-1\right)^+,\\ &R_2 \leq \left(\frac{1}{2}\log\left(h_{22}\right)-1\right)^+,\\ &R_1+R_2 \leq \left(\frac{1}{2}\log\left(\mu_g\right)+\frac{1}{2}\log\left(\lambda_h\right)-\frac{3}{2}\right)^+,\\ &R_1+R_2 \leq \left(\frac{1}{2}\log\left(\mu_h\right)+\frac{1}{2}\log\left(\lambda_g\right)-\frac{3}{2}\right)^+\Big\}.\end{aligned}$$

Note that the RHS's of the sum-rate bounds depend on the order of the channel gains. For most of possible orderings, these two inequalities would be consequences of the individual rate bounds. For example, if $\lambda_h = h_{22}$, then the last bound is implied by the first and fourth bounds, since $\mu_g \geq g_{11}$. It can be shown in general that $\mathcal{R}_{\text{ach}}^{\mathsf{GZZ}}$ is equivalent to

$$\begin{aligned}\mathcal{R}_{\text{ach}}^{\mathsf{GZZ}} = \Big\{(R_1, R_2) : & R_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \\ & R_1 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_g) + \frac{1}{2}\log(h_{12}) - \frac{3}{2}\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_h) + \frac{1}{2}\log(g_{12}) - \frac{3}{2}\right)^+ \Big\}.\end{aligned}$$

Now, note that $g_{11} \geq 1$, $g_{12} \geq 1$, and $g_{22} \geq 1$. These imply

$$\begin{aligned}\frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) &\leq \frac{1}{2}\log(3\max\{g_{11}, g_{12}\}) + \frac{1}{2}\log\left(\frac{2\max\{g_{12}, g_{22}\}}{g_{12}}\right) \\ &\leq \frac{1}{2}\log\left(\frac{\max\{g_{11}, g_{12}\} \cdot \max\{g_{22}, g_{12}\}}{g_{12}}\right) + \frac{1}{2}\log 6 \\ &\leq \frac{1}{2}\log(\mu_{h,\mathsf{Z}}) + \frac{1}{2}\log 6. \qquad \text{(B.38)}\end{aligned}$$

We also have

$$\frac{1}{2}\log(1 + x) \leq \frac{1}{2}\log(x) + \frac{1}{2} \qquad \text{(B.39)}$$

for all $x \geq 1$. Applying (B.38) and (B.39), we obtain the following achievable rate region, which is a

subset of $\mathcal{R}_{\text{ach}}^{\mathsf{GZZ}}$.

$$\begin{aligned}\mathcal{R}_{\text{ach,2}}^{\mathsf{GZZ}} = \Big\{(R_1,R_2) : & R_1 \le \left(\frac{1}{2}\log(1+g_{11})-\frac{3}{2}\right)^+,\\ & R_2 \le \left(\frac{1}{2}\log(1+g_{22})-\frac{3}{2}\right)^+,\\ & R_1 \le \left(\frac{1}{2}\log(1+h_{11})-\frac{3}{2}\right)^+,\\ & R_2 \le \left(\frac{1}{2}\log(1+h_{22})-\frac{3}{2}\right)^+,\\ & R_1+R_2 \le \left(\frac{1}{2}\log(1+g_{11}+g_{12})+\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right)+\frac{1}{2}\log(1+h_{12})-\frac{7}{2}\right)^+,\\ & R_1+R_2 \le \left(\frac{1}{2}\log(1+h_{11}+h_{12})+\frac{1}{2}\log\left(1+\frac{h_{22}}{h_{12}}\right)+\frac{1}{2}\log(1+g_{12})-\frac{7}{2}\right)^+\Big\}\end{aligned}$$

Therefore, for any rate pair $(R_1,R_2) \in \mathcal{R}^{\mathsf{GZZ}}$, the rate pair $(R_1 - \frac{1}{4}\log 12, R_2 - \frac{1}{4}\log 12)$ belongs to $\mathcal{R}_{\text{ach,2}}^{\mathsf{GZZ}}$, and therefore can be achieved using the proposed encoding scheme.

APPENDIX C
PROOF OF LEMMAS

Discussion of Example 4 in Section III: The converse proof is fairly simple and follows from a similar argument we used to prove (GZS-1), (GZS-2), and (GZS-3) in Appendix A.

In the following we will present an encoding strategy which guarantees to achieve rate pair $(R_1 - \frac{1}{2}, R_2 - \frac{1}{2})$, provided that $(R_1,R_2) \in \mathcal{R}^{\mathsf{Z}}$. This gives us an approximate capacity characterization for the Gaussian Z network. In order to do this, we consider the following two cases.

Case A: $g_{12} \ge g_{22}$

Assume (R_1,R_2) be an achievable rate pair. Then, the first receiver G_1 is able to decode W_1 sent at rate R_1, and remove the signal associated to W_1 from its received signal. The remaining signal provides a higher SNR to decode W_2 than the signal received at G_2. Therefore, in this particular regime, the first receive would be able to decode both messages. Hence, we have a Gaussian multiple access channel from F_1 and F_2 to G_1, combined with a line network from F_2 to G_2. Therefore, the intersection of the

rate regions of the Gaussian MAC and the line networks is simply achievable. That is

$$\begin{aligned}\mathcal{R}^{\mathsf{Z}}_{\text{ach,A}} =& \left\{(R_1,R_2): R_1 \le \frac{1}{2}\log\left(1+g_{11}\right), R_2 \le \frac{1}{2}\log\left(1+g_{12}\right), R_1+R_2 \le \frac{1}{2}\log\left(1+g_{11}+g_{12}\right)\right\} \\ & \bigcap \\ & \left\{(R_1,R_2): R_2 \le \frac{1}{2}\log\left(1+g_{22}\right)\right\} \\ =& \left\{(R_1,R_2): R_1 \le \frac{1}{2}\log\left(1+g_{11}\right), R_2 \le \frac{1}{2}\log\left(1+g_{22}\right), R_1+R_2 \le \frac{1}{2}\log\left(1+g_{11}+g_{12}\right)\right\}.\end{aligned} \tag{C.1}$$

Note that the individual rate bounds in $\mathcal{R}^{\mathsf{Z}}$ and $\mathcal{R}^{\mathsf{Z}}_{\text{ach,A}}$ are the same. Moreover, the difference between the sum rate bounds is bounded by

$$\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) \le \frac{1}{2}\log\left(1+1\right) = \frac{1}{2}. \tag{C.2}$$

Therefore, the gap between each boundary point of $\mathcal{R}^{\mathsf{Z}}$ and $\mathcal{R}^{\mathsf{Z}}_{\text{ach,A}}$ is at most $\frac{1}{2}$ bit.

Case B: $g_{12} \le g_{22}$*:*

The encoding scheme we introduce for this case is similar to Han-Kobayashi's scheme for 2-user interference channel. We first split the second message W_2 into the common and private parts, $W_2 = (W_2^c, W_2^p)$, with rates R_2^c and R_2^p, respectively, where W_2^c can be decoded at both receivers and W_2^p is only decodable at G_2. Sub-messages W_1, W_2^c, and W_2^p are encoded by corresponding randomly generated Gaussian codes to $\mathbf{x}_1$, $\mathbf{x}_2^c$ and $\mathbf{x}_2^p$, and the resulting codewords are sent over the channel.

We allocate $\alpha_p = 1/g_{12}$ fraction of the transmission power available at F_2 to W_2^p, and the remaining power $\alpha_c = 1-\alpha_p$ is allocated to W_2^c. Therefore, we have

$$\mathbf{x}_2 = \sqrt{\alpha_c}\mathbf{x}_2^p + \sqrt{\alpha_c}\mathbf{x}_2^p.$$

The first receiver, G_1, decodes W_1 and W_2^c treating W_2^p as noise. Therefore, the effective noise power received at G_1 would be $\mathbb{E}[\sqrt{g_{12}\alpha_p}x_p + z_1]^2 = 2$. According to the capacity region of Gaussian multiple access channel, this can be done provided that

$$\begin{aligned} R_1 &\le \tfrac{1}{2}\log\left(1+\tfrac{g_{11}}{2}\right), \\ R_2^c &\le \tfrac{1}{2}\log\left(\tfrac{1+g_{12}}{2}\right), \\ R_1+R_2^c &\le \tfrac{1}{2}\log\left(\tfrac{1+g_{11}+g_{12}}{2}\right). \end{aligned} \tag{C.3}$$

The second decoder first decodes W_2^c treating W_2^p as noise. It then removes the corresponding codeword from the received signal, and decodes W_2^p. This can be done as long as

$$\begin{aligned} R_2^c &\le \tfrac{1}{2}\log\left(\tfrac{1+g_{22}}{1+g_{22}/g_{12}}\right), \\ R_2^p &\le \tfrac{1}{2}\log\left(1+\tfrac{g_{22}}{g_{12}}\right). \end{aligned} \tag{C.4}$$

Note that we have two upper bounds for R_2^c. However, it is easy to show that $\frac{1+g_{22}}{1+g_{22}/g_{12}} \geq \frac{1+g_{12}}{2}$, for $1 \leq g_{12} \leq g_{22}$, and therefore, the first bound dominates the second one. Using Fourier-Motzkin elimination to write the achievable region in terms of R_1 and $R_2 = R_2^c + R_2^p$, and after some simplification, we get that the region

$$\mathcal{R}_{\text{ach,B}}^{\mathsf{Z}} = \Big\{(R_1, R_2) : R_1 \leq \frac{1}{2}\log(1+g_{11}) - \frac{1}{2}, \tag{C.5}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{22}) - \frac{1}{2}, \tag{C.6}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}.\Big\}. \tag{C.7}$$

is achievable. Therefore, if $(R_1, R_2) \in \mathcal{R}^{\mathsf{Z}}$, then $(R_1 - \frac{1}{2}, R_2 - \frac{1}{2})$ is achievable. ■

Proof of Lemma 1 in Section V: The following achievability scheme simply uses superposition encoding of sub-messages at F_2, and a successively decode and cancel strategy at G_1 and G_2. We use a random codebook with a proper number of codewords, generated according to a zero-mean unit-variance Gaussian distribution for each message. A proper power allocation for the messages at the transmitters allow the decoders to apply a decode and cancel strategy. We denote the codeword corresponding to the message $U_i^{(j)}$ by $\mathbf{x}_{i,j}$, and the power allocated to this message by $\alpha_{i,j}$.

The available power at F_2 can be arbitrarily allocated to its sub-messages. In particular, we choose the power coefficients so that they satisfy $\alpha_{2,2} \leq 1/g_{22}$, $\alpha_{2,3} \leq 1/g_{12}$, and $\alpha_{2,1} = 1 - \alpha_{2,2} - \alpha_{2,3}$. In the decoding part, G_1 and G_2 treat $U_2^{(3)}$ and $U_2^{(2)}$, respectively, as noise. Therefore, the total noise at G_1 and G_2 would be $\tilde{\mathbf{z}}_1 = \sqrt{g_{12}\alpha_{2,3}}\mathbf{x}_{2,3} + \mathbf{z}_1$ and $\tilde{\mathbf{z}}_2 = \sqrt{g_{22}\alpha_{2,2}}\mathbf{x}_{2,2} + \mathbf{z}_2$. However, the effective noise power cannot exceed 2 since $\mathbb{E}[g_{12}\alpha_{2,3} + 1] \leq 2$ and $\mathbb{E}[g_{22}\alpha_{2,2} + 1] \leq 2$.

The receiver F_1 observes a Gaussian multiple access channel (with noise power upper bounded by 2), where $U_1^{(1)}$ is sent by one user, and $(U_2^{(1)}, U_2^{(2)})$ is sent by the other user. The bounds in (7)-(10) guarantee that these rates are achievable over the multiple access channel.

On the other hand, the channel from F_2 to G_2 is Gaussian point-to-point channel with modified additive noise. Therefore, any total rate not exceeding its capacity can be reliably transmitted. This is condition is fulfilled here since $\Upsilon_{2,1} + \Upsilon_{2,3}$ satisfies (12). Finally, the bound on the power allocated to $U_2^{(3)}$ upper bounds its rate as in (11).

■

Proof of Lemma 2 in Section V: Again, the achievability scheme we propose for the Gaussian S interference network (illustrated in Figure 11) is based on superposition coding, and a successively decode and cancel decoding strategy, such that the requirements of the problem are fulfilled. A proper

power allocation is required to guarantee achievability of the rate tuples mentioned in this lemma.

Note that G_1 does not decode $V_2^{(2)}$ and $V_2^{(4)}$, and treats them as noise. We choose the total fraction of power allocated to $V_2^{(2)}$ and $V_2^{(4)}$ to be at most $1/h_{11}$, that is $\alpha_{2,2}+\alpha_{2,4}\leq 1/h_{11}$. Therefore, the total noise power received at G_1 is upper bounded as $\mathbb{E}[h_{11}(\alpha_{2,2}+\alpha_{2,4})+1]\leq 2$.

Similarly, $V_1^{(2)}$ is treated as noise at G_2. By bounding the fraction of power allocated to this sub-message, we can upper bound the effective noise power observed at G_2 by $\mathbb{E}[h_{11}(\alpha_{2,2}+\alpha_{2,4})+1]\leq 2$.

The point-to-point Gaussian channel from F_1 to G_1 can support any sum-rate below its capacity as in (13). Moreover, $\Theta_{1,2}$ is bounded above since its allocated power does not exceed $1/h_{12}$.

On the other hand, we have a Gaussian multiple access channel from F_1 and F_2 to G_2, with total noise power not exceeding 2. The bounds in (15)-(18) guarantee that the desired rates belong to the capacity region of this channel, and therefore they are achievable. We skip the details of power allocation here, but we point out that the achievability of the region is a consequence of the Gaussian multiple access rate region achievability.

■

Proof of Lemma 3 in Section V: In this part we show that any rate tuple satisfying (21)-(24) is achievable. The main idea of this proof can be summarized as follows.

- Use a common codebook with group structure, such as lattice codes, for $W_1^{(0)}$ and $W_2^{(0)}$, which maps them to $\mathbf{x}_{1,0}$ and $\mathbf{x}_{2,0}$
- Choose a proper power allocation for $\mathbf{x}_{1,0}$ and $\mathbf{x}_{2,0}$ such that they get received at G_1 at the same power level; More precisely, denoting their power allocation by α_0 and β_0, they should satisfy $g_{11}\alpha_0 = g_{12}\beta_0$. This condition guarantees that the two lattice points get scaled by the same factor, and therefore the result is still a lattice point on the scaled lattice and can be decoded as long as enough signal to noise ratio is provided.
- Use random Gaussian codebooks to encode the private sub-messages to $\mathbf{x}_{1,1}$ and $\mathbf{x}_{1,2}$, and use proper power allocation, α_1 and β_1.

The first receiver G_1 needs to decode the partial-invertible ϕ which we define as

$$\begin{aligned}\phi\left(W_1^{(0)},W_2^{(0)}\right) &= \psi^{-1}\left(\psi\left(W_1^{(0)}\right)+\psi\left(W_1^{(0)}\right)\right)\\ &= \psi^{-1}\left(\mathbf{x}_{1,0}+\mathbf{x}_{2,0}\right)\end{aligned}$$

where ψ is the one-to-one encoding function which maps the functional messages to the common lattice codebook. Note that the group structure of the code impels that $\mathbf{x}_{1,0}+\mathbf{x}_{2,0}$ is still a valid codeword. It is easy to check that this function is partial-invertible.

Let us define

$$\eta = \min\left\{g_{11}, g_{12}, g_{22}, \frac{g_{11}g_{22}}{g_{12}}\right\}. \tag{C.8}$$

Depending on the minimizer in η, we identify four cases. In each case, the achievable rate region is a polytope, with a certain number of corner points. It suffices to show the achievability only for the cornet points, since a standard time-sharing argument guarantees achievability for the rest of the region.

The proof details for each corner point includes message splitting, and power allocation for sub-messages such that the decoders be able to decode corresponding messages. In the following we describe this strategy in details for the case where $\eta = g_{11}$. The extension of this method for other cases is straight-forward, and therefore we skip it here to sake of brevity.

Case I. $\eta = g_{11}$*:* It is clear from the definition of η that in this case $g_{11} \leq g_{12} \leq g_{22}$, and therefore $\lambda = g_{11}$ and $\mu = g_{22}$. Hence, the desired region is characterized by all non-negative rate tuples (P_0, P_1, P_2) satisfying

$$P_0 + P_1 \leq \frac{1}{2}\log\left(g_{11}\right),$$
$$P_0 + P_1 + P_2 \leq \frac{1}{2}\log\left(g_{22}\right).$$

This rate region is illustrated in Figure 18. It suffices to show that the corner points A, B and C are achievable, since the points D and E are degenerated from B and C, respectively.

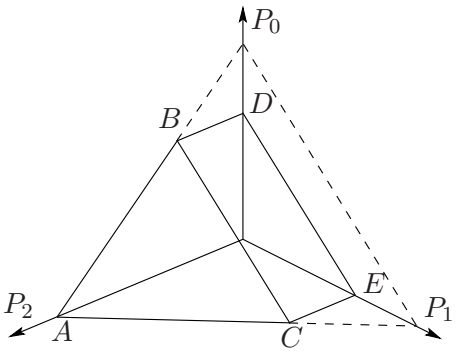


Fig. 18. Achievable rate region of the Z-neutralization network when $\eta = g_{11}$.

- $A : (P_0, P_1, P_2) = \left(0, 0, \frac{1}{2}\log\left(g_{22}\right) - \frac{3}{2}\right)$

  The encoding strategy for this corner point is fairly simple. The second transmitter uses all its available power to send $W_2^{(1)}$, while the first transmitter keeps silent. That is, $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = \mathbf{x}_{2,1}$.

The first decoder has nothing to decode, and the second one can decode $\mathbf{x}_2$ from $\mathbf{y}_2$ as long as $P_2 \leq \frac{1}{2}\log(1+g_{22})$. It is clear that in particular $P_2 = \frac{1}{2}\log(g_{22}) - \frac{3}{2}$ is achievable.

- $B : (P_0, P_1, P_2) = \left(\frac{1}{2}\log(g_{11}) - 1, 0, \frac{1}{2}\log(g_{22}) - \frac{1}{2}\log(g_{11}) - \frac{1}{2}\right)$

  The first encoder sends its lattice codeword with power allocation $\alpha_0 = (g_{11}-1)/g_{11}$. The second encoder splits its private message into $W_2^{(1)} = (W_2^{(1,1)}, W_2^{(1,2)})$ of rates $P_{2,1}$ and $P_{2,2}$ where $P_2 = P_{2,1} + P_{2,2}$. Then it sends

$$\mathbf{x}_2 = \sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + \sqrt{\beta_0}\mathbf{x}_{2,0} + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}$$

  where the power allocation coefficients are fixed to be $\beta_{1,2} = 1/g_{12}$, $\beta_0 = (g_{11}-1)/g_{12}$, and $\beta_{1,1} = 1 - \beta_0 - \beta_{1,2}$. The signal received at the destinations are

$$\begin{aligned}\mathbf{y}_1 &= \sqrt{g_{11}}\mathbf{x}_1 + \sqrt{g_{12}}\mathbf{x}_2 + \mathbf{z}_1,\\ &= \sqrt{g_{12}-g_{11}}\mathbf{x}_{2,1,1} + \sqrt{g_{11}-1}[\mathbf{x}_{1,0}+\mathbf{x}_{2,0}] + \mathbf{x}_{2,1,2} + \mathbf{z}_1, \qquad \text{(C.9)}\\ \mathbf{y}_2 &= \sqrt{g_{22}}\mathbf{x}_2 + \mathbf{z}_2,\\ &= \sqrt{\frac{g_{22}(g_{12}-g_{11})}{g_{12}}}\mathbf{x}_{2,1,1} + \sqrt{\frac{g_{22}(g_{11}-1)}{g_{12}}}\mathbf{x}_{2,0} + \sqrt{\frac{g_{22}}{g_{12}}}\mathbf{x}_{2,1,2} + \mathbf{z}_2. \qquad \text{(C.10)}\end{aligned}$$

  The first node decode and cancel $\mathbf{x}_{2,1,1}$, $\tilde{\mathbf{x}}_0 = \mathbf{x}_{1,0} + \mathbf{x}_{2,0}$, and $\mathbf{x}_{2,1,2}$ in order, while the second one performs the same decoding for $\mathbf{x}_{2,1,1}$, $\mathbf{x}_{2,0}$, and $\mathbf{x}_{2,1,2}$. It is easy to show that the rates $P_{2,1} = \frac{1}{2}\log(g_{12}/g_{11}) - 0.5$, $P_0 = \frac{1}{2}\log(g_{11}) - 1$, and $P_{2,2} = \frac{1}{2}\log(g_{22}/g_{12})$ are achievable, which implies the private rates $P_2 = P_{2,1} + P_{2,2} = \frac{1}{2}\log(g_{22}/g_{11}) - \frac{1}{2}$ for the second transmitter.

- $C : (P_0, P_1, P_2) = \left(0, \frac{1}{2}\log(g_{11}) - 1, \frac{1}{2}\log(g_{22}) - \frac{1}{2}\log(g_{11}) - \frac{1}{2}\right)$

  For this rate tuple, the rate of the functional message is zero. The second transmitter splits its private message similar to that of corner point B. The transmission power is distributed between among the sub-message as $\alpha_0 = 0$, $\alpha_1 = 1$, $\beta_{1,2} = 1/g_{12}$, $\beta_0 = 0$, and $\beta_{1,1} = 1 - \beta_{1,2}$. A similar argument to that of corner point B shows that the rates $P_{2,1} = \frac{1}{2}\log(g_{12}/g_{11}) - 0.5$, $P_1 = \frac{1}{2}\log(g_{11}) - 1$, and $P_{2,2} = \frac{1}{2}\log(g_{22}/g_{12})$ are achievable, which implies the achievability of the rate point C.

■

Proof of Lemma 4 in Section VI.2: Let $(R_1, R_2) \in \mathcal{R}^{\mathsf{DZS}}$ be an arbitrary rate pair which satisfies (DZS-1)-(DZS-10). In particular $R_1 \leq \min\{m_{11}, n_{11}\}$. We claim that $(R_1, R_2) \in \mathcal{R}_1^{\mathsf{DZS}}(t) \times \mathcal{R}_2^{\mathsf{DZS}}(t)$ for $t = R_1$, and therefore (R_1, R_2) is achievable using network decomposition. In order to do this we have to show that any R_2 satisfying (DZS-1)-(DZS-10), fulfills the constraints in the definition of $\mathcal{R}_2^{\mathsf{DZS}}(R_1)$.

Using (DZS-2) and (DZS-3), we have

$$\begin{aligned} R_2 &\le \min\left(\max(m_{11},m_{12}) + (m_{22}-m_{12})^+ - R_1, \max(m_{12},m_{22})\right) \\ &= \min\left(\max(m_{11},m_{12}) - R_1, m_{12}\right) + (m_{22}-m_{12})^+ \\ &= m'_{12}(R_1) + (m_{22}-m_{12})^+ \\ &\le m'_{12}(R_1) + (m'_{22}(R_1) - m'_{12}(R_1))^+ \qquad \text{(C.11)} \\ &= \max(m'_{12}(R_1), m'_{22}(R_1)), \qquad \text{(C.12)} \end{aligned}$$

where in (C.11) we have used the fact that

$$(\min(a,b) - \min(c,d))^+ \ge \min\left((a-c)^+, (b-d)^+\right).$$

Moreover, since R_2 satisfies (DZS-3), (DZS-5), and (DZS-8), we have

$$\begin{aligned} R_2 &\le \min\left(\max(m_{11},m_{12}) + (m_{22}-m_{12})^+ - R_1, m_{22} + \max(n_{11},n_{21}) - R_1, m_{22}+n_{21}\right) \\ &\le \min\left(\max(m_{11},m_{12}) + (m_{22}-m_{12})^+ - R_1, m_{22}\right) + \min\left(\max(n_{11},n_{21}) - R_1, n_{21}\right) \qquad \text{(C.13)} \\ &= m'_{22}(R_1) + n'_{21}(R_1), \qquad \text{(C.14)} \end{aligned}$$

where (C.13) holds since

$$\min(a,b) + \min(c,d) \ge \min(a, b+c, b+d),$$

for non-negative a, b, c, and d.

In order to show that the third constraint is satisfied, we can start with (DZS-4), (DZS-6), and (DZS-10).

$$\begin{aligned} R_2 &\le \min\left(\max(m_{11},m_{12}) + n_{22} + -R_1, m_{12}+n_{22}, \max(n_{11},n_{21}) + (n_{22}-n_{21})^+ - R_1\right) \\ &\le \min\left(\max(m_{11},m_{12}) + -R_1, m_{12}\right) + \min\left(\max(n_{11},n_{21}) + (n_{22}-n_{21})^+ - R_1, n_{22}\right) \\ &= m'_{12}(r_1) + n'_{22}(r_1). \qquad \text{(C.15)} \end{aligned}$$

Finally, using (DZS-8) and (DZS-10), we have

$$\begin{aligned} R_2 &\le \min\left(\max(n_{11},n_{21}) + (n_{22}-n_{21})^+ - r_1, \max(n_{21},n_{22})\right) \\ &= \min\left(\max(n_{11},n_{21}) - r_1, n_{21}\right) + (n_{22}-n_{21})^+ \\ &= n'_{21}(r_1) + (n_{22}-n_{21})^+ \\ &\le n'_{21}(r_1) + (n'_{22}(r_1) - n'_{21}(r_1))^+ \\ &= \max(n'_{21}(r_1), n'_{22}(r_1)). \qquad \text{(C.16)} \end{aligned}$$

Putting inequalities in (C.12) and (C.14)-(C.16) together shows that $R_2 \in \mathcal{R}_2^{\mathsf{DZS}}(R_1)$, and completes the proof. ■

Proof of Lemma 5 in Section VII.2:

The coding strategy we present here is based a network decomposition, where the sub-nodes and the links of the deterministic Z-interference network are partitioned into two disjoint sets. We analyze the rate region of each network, and derive an achievable rate region for the original network based on this analysis.

We just point out here that in this coding strategy, the second sender F_2, never sends a bit on a sub-node which is not received at G_2, even if $n_{12} > n_{22}$.

The first partition of the network $\mathcal{N}_1$, consists of those sub-nodes in G_1 which are connected to one of the top m_{11} sub-nodes of F_1 *and* one of the top m_{22} sub-nodes of F_2. All the sub-nodes in the network which are related to (see Definition 3) any of these sub-nodes also belong to the first network partition. The remaining nodes and link form the second part of the network $\mathcal{N}_2$. It is clear that these two networks are node-disjoint, and do not cause interference on each other.

We first characterize the number sub-nodes in G_1 which belong to $\mathcal{N}_1$, by determining whether each of them can receive a bit from F_1, F_2, or both of them. We denote the number of levels in G_1 which are only connected to a transmitting level in F_1 by k_1. Similarly, the number of those only connected to a a transmitting level (the top $\min(n_{12}, n_{22})$) in F_2 by k_2. Finally, k_0 denotes the number of levels which are connected to transmitting levels of both F_1 and F_2 (see Figure 19).

First, we derive k_0. Enumerate the levels of G_1 from 1 (for the highest) to q (for the lowest). Let j be the index of a sub-node in G_1 belong to $\mathcal{N}_1$, *i.e.,* it receives bits from both F_1 and F_2. Its neighbors in F_1 and F_2 (if there is any) are indexed by $j + n_{11} - q$ and $j + n_{12} - q$, respectively. Therefore, j belongs to $\mathcal{N}_1$ if and only if $1 \leq j + n_{11} - q \leq n_{11}$ and $1 \leq j + n_{12} - q \leq \min(n_{12}, n_{22})$. Therefore, the number of such sub-nodes is given by

$$\begin{aligned} k_0 &= [\min\{q, q - n_{12} + n_{22}\} - \max\{q - n_{11}, q - n_{12}\}]^+ \\ &= \min\{n_{11}, n_{12}, n_{22}, (n_{11} + n_{22} - n_{12})^+\}. \end{aligned} \tag{C.17}$$

It is clear from the definition of k_0 that the remaining $n_{11} - k_0$ lowest levels of G_1 are only connected to sub-nodes of F_1, and hence, $k_1 = n_{11} - k_0$. Similarly, $\min\{n_{12}, n_{22}\}$ sub-nodes in G_1 are receiving information from F_2, where k_0 of them are also connected to F_1. Therefore, the remaining sub-nodes are only connected to G_2. Thus, $k_2 = \min\{n_{12}, n_{22}\} - k_0$.

We partition the network into two parts: The first part consists of the k_0 sub-nodes of G_1 connected to

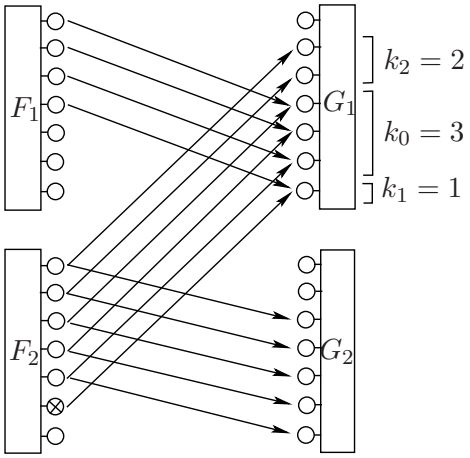


Fig. 19. A deterministic Z-neutralization network. The upper 2 sub-nodes in G_1 are only connected to F_2, and therefore $k_2 = 2$. The next 3 sub-nodes receive information from both F_1 and F_2, and hence $k_0 = 3$. Although the lowest sub-node is also connected to both transmitters, it only receives information from F_1 since F_2 keeps silent on its sub-nodes below n_{22}.

both F_1 and F_2, and sub-nodes connected to them. The remaining sub-nodes form the second partition of the network. We characterize the achievable tuples for each, denoted by (Q_0', Q_1', Q_2') and (Q_0'', Q_1'', Q_2''), respectively. The fact that these two partitions are isolated allows us to conclude that the summation of such achievable tuples is also achievable for the original network.

Consider the first partition of the network. It is clear that any of the k_0 levels of G_1 connected to both F_1 and F_2 and can be used to communicate a functional bit, since G_1 naturally receives the `xor` of the transmitting bits. On the other hand, such sub-node can be used to communicate one private bit from any of F_1 or F_2 to G_1 by keeping the other one silent. Therefore, any rate tuple satisfying

$$Q_0' + Q_1' + Q_2' \leq k_0 \tag{C.18}$$

is achievable.

The non-interfered links of the second partition of the network can be used to send private bits from the transmitters to G_1 simultaneously. Moreover, each transmitter can use one of its non-interfering sub-nodes to send a functional bit to G_1, and then, G_1 computes their `xor`, after receiving them separately. This can provide up to $\min\{k_1, k_2\}$ new functional bits for G_1. Moreover, the lower $(n_{22} - n_{12})^+$ sub-nodes of F_2 which are connected to G_2 but not to G_1 can be used to send private bits to G_2 without causing any interference at G_1.

Hence, this strategy can transmit any rate tuple satisfying

$$\begin{aligned} Q_0'' &\leq \min\{k_1, k_2\}, \\ Q_0'' + Q_1'' &\leq k_1, \\ Q_0'' + Q_2'' &\leq k_2 + (n_{22} - n_{12})^+. \end{aligned} \tag{C.19}$$

Summing up the rates achieved on each partition of the network, we have arrive at $Q_i = Q_i' + Q_i''$ for $i = 0, 1, 2$, where (Q_0', Q_1', Q_2')'s and (Q_0'', Q_1'', Q_2'') satisfy (C.18) and (C.19), respectively. It only remains to apply the Fourier-Motzkin elimination to project the rate region on the (Q_0, Q_1, Q_2) space. This gives us

$$\begin{aligned} Q_0 &\leq k_0 + \min\{k_1, k_2\}, \\ Q_0 + Q_1 &\leq k_0 + k_1, \\ Q_0 + Q_2 &\leq k_0 + k_2 + (n_{22} - n_{12})^+, \\ Q_0 + Q_1 + Q_2 &\leq k_0 + k_1 + k_2 + (n_{22} - n_{12})^+. \end{aligned} \tag{C.20}$$

Some simple manipulations show that the RHS's of the inequalities in (C.20) are the same as that claimed in the lemma.

■

Proof of Lemma 6 in Appendix A.1: As mentioned before, we will use the Fano's inequality in order to prove this lemma. We have

$$\begin{aligned} \ell R_1 = H(W_1) &= I(W_1; y_1^\ell) + H(W_1|y_1^\ell) \\ &\leq I(W_1; y_1^\ell) + \ell e_\ell \qquad \text{(C.21)} \\ &\leq I(x_1^\ell; y_1^\ell) + \ell e_\ell, \qquad \text{(C.22)} \end{aligned}$$

where (C.21) is implied by the Fano's inequality, and in (C.22) we used the data processing inequality for the Markov chain $W_1 \leftrightarrow x_1^\ell \leftrightarrow y_1^\ell$. Note that where $\varepsilon_\ell \to 0$ as ℓ grows. The proofs of the other two inequalities follow the same lines, and we skip them to sake of brevity. ■

Proof of Lemma 7 in Appendix B.1: Note that Z is independent of everything else, and X_1 and X_2 are conditionally independent. Without loss of generality we can also assume that $\mu_i(\gamma) = \mathbb{E}[X_i|\Gamma = \gamma] = 0$ for $\forall\gamma$ (otherwise for any given $\Gamma = \gamma$, we can shift X_i by $\mu_i(\gamma)$, while the entropy does not change). Let $\mathbb{E}[X_i^2|\Gamma = \gamma] = \sigma_i^2(\gamma)$ for $i = 1, 2$. Therefore the conditional variance of Y can be bounded as

$$\mathbb{E}[Y^2|\Gamma = \gamma] = \mathbb{E}[(X_1 + X_2 + Z)^2|\Gamma = \gamma] = \sigma_1^2(\gamma) + \sigma_2^2(\gamma) + 1. \tag{C.23}$$

Therefore,

$$h(Y|\Gamma) = \mathbb{E}_\Gamma[h(Y|\Gamma=\gamma)] = \mathbb{E}_T[h(X_1+X_2+Z|\Gamma=\gamma)]$$
$$\leq \mathbb{E}_\Gamma[\log 2\pi e(\sigma_1^2(\gamma)+\sigma_2^2(\gamma)+1)] \quad \text{(C.24)}$$
$$\leq \log 2\pi e(\mathbb{E}_\Gamma[\sigma_1^2(\gamma)+\sigma_2^2(\gamma)+1]) \quad \text{(C.25)}$$
$$= \log 2\pi e(\sigma_1^2+\sigma_2^2+1), \quad \text{(C.26)}$$

where in (C.24) we have used the fact that Gaussian random variable has the maximum differential entropy among all random variables with the same variance, and (C.25) follows from the concavity of the function $\log(\cdot)$. Finally, (C.26) is just the tower property, $\mathbb{E}_\Gamma[\mathbb{E}[X_i^2|\Gamma]] = \mathbb{E}[X_i^2]$. ■

REFERENCES

[1] A. Schrijver, *Theory of Linear and Integer Programming*. New York: Wiley, 1998.

[2] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, pp. 49–60, January 1981.

[3] R. H. Etkin, D. Tse, , and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[4] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[5] K. Gomadam and S. A. Jafar, "The effect of noise correlation in amplifyand-forward relay networks," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 731–745, Feb. 2009.

[6] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[7] A. Ozgur, O. Leveque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[8] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3413–3424, Aug. 2009.

[9] V. Prabhakaran and P. Viswanath, "Interference channel with destination cooperation," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

[10] C. Suh and D. Tse, "Symmetric feedback capacity of the gaussian interference channel to within one bit," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

[11] G. Bresler and D. Tse, "The Two-User Gaussian Interference Channel: A Deterministic View," *European Transactions in Telecommunications*, vol. 19, pp. 333–354, June 2008.

[12] V. Cadambe and S. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Trans. Information Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008, .

[13] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication Over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE Trans. Information Theory*, vol. 54, no. 8, pp. 3457–3470, August 2008.

[14] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[15] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, "Transmission techniques for relay-interference networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2008.

[16] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[17] S. Mohajer, S. Diggavi, C. Fragouli, and D. N. C. Tse, "Capacity of deterministic z-chain relay-interference network," in *Proceedings of IEEE Information Theory Workshop*, Volos, Greece, 2009, pp. 331–335.