Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Capacity Results for a Class of Z Channels with Degraded Message Sets

Nan Liu and Wei Kang

***Abstract*— We study a two-transmitter two-receiver network where Receiver 1 can only hear the transmitted signal of Transmitter 1. Transmitter 1 has two messages, one of which is intended for Receiver 1 while both are intended for Receiver 2. Transmitter 2 has one message which is intended for Receiver 2. We call this channel model the Z channel with degraded message sets. When the multiple access link between the two transmitters and Receiver 2 satisfies certain conditions, we characterize the capacity region of the Z channel with degraded message sets despite the presence of distributed encoding.**

I. INTRODUCTION

The Z channel, see Figure 1, is a two-transmitter two-receiver network where Receiver 2 can hear the transmitted signals of both transmitters and Receiver 1 can only hear the transmitted signal of Transmitter 1. Transmitter 1 has two messages, W_{11} and W_{12}. Message W_{11} is intended for Receiver 1, and message W_{12} is intended for Receiver 2. Transmitter 2 has one message W_2 which is intended for Receiver 2. All messages are independent. The Z channel was first introduced by Vishwanath et al. in [1]. Since then, researchers have been trying to find the capacity region of the Z channel [2]–[8]. The Z channel contains the broadcast channel (setting the rate of W_2 to be zero) and the Z-interference channel (setting the rate of W_{12} to be zero) as special cases. Since the capacity regions of the general broadcast channel and the general Z-interference channel are both open, finding the capacity region of the Z channel is difficult.

While the capacity region of the general broadcast channel is still open, the capacity region of the broadcast channel with degraded message sets is completely known [9]. Modifying the channel model of the Z channel to include the broadcast channel with degraded message sets rather than the broadcast channel, there are two possibilties. Figure 2 depicts the first possibility: Transmitter 1 has two messages W_{1c} and W_{1p}. Message W_{1c} is intended for Receiver 2 while both messages W_{1c} and W_{1p} are intended for Receiver 1. This channel model was mentioned in [10]. Figure 3 depicts the second possibility: Transmitter 1 has two messages W_{1c} and W_{1p}. Message W_{1c} is intended for Receiver 1 while both messages W_{1c} and W_{1p} are intended for Receiver 2. Comparing the two possibilities of Figures 2 and 3, we find that in Figure 2, when we set the rate of message W_{1c} to be zero, we obtain the Z-interference channel, which means that the channel model in Figure 2 contains the Z-interference channel as a special case and therefore, finding the capacity region of the channel model in Figure 2 is at least as hard as that of the Z-interference channel. The channel model in Figure 3 however does not include the Z-interference channel as a special case. From this perspective, finding the capacity region of the channel model in Figure 3 is probably easier than the Z channel and the channel model in Figure 2. This is why we focus on the channel model in Figure 3 in this paper. We call this the Z channel with degraded message sets.

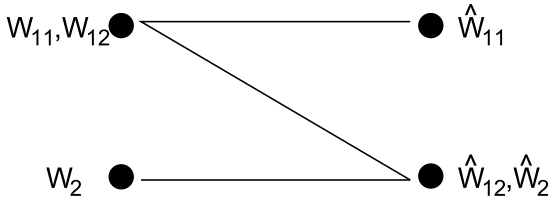

Fig. 1. The Z channel.

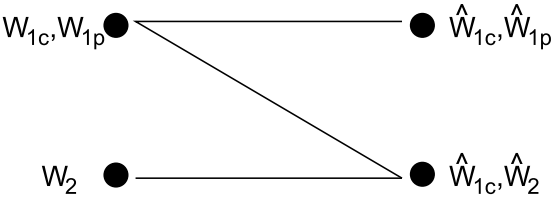

Fig. 2. Possibility 1.

We first provide a general achievability scheme for the Z channel with degraded message sets. The scheme uses superposition encoding at Transmitter 1 and combines the achievability techniques of the broadcast channel with degraded message sets and the multiple access channel. We then go on to show that this achievability scheme is optimal for the Z channel with degraded message sets when the multiple access link between the two transmitters and Receiver 2 satisfies certain conditions. The converse techniques that we use include the introduction of imaginary channels [11], the single-letterization technique [12, page 314], and the technique of replacing two auxiliary random variables with one [13].

The Z channel with degraded message sets in Figure 3 contains the element of distributed encoding which is also contained in traditional channel models like the interference channel. Therefore, solving the Z channel with degraded message sets helps with the understanding of how to deal

This work is partially supported by the National Natural Science Foundation of China under Grants 61201170 and 61271208, China-EU International Scientific and Technological Cooperation Program (0902), International Science and Technology Cooperation Program under Grant 2008DFA12090, and the New Teacher Funds of Southeast University.

N. Liu is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China nanliu@seu.edu.cn

W. Kang is with the School of Information Science and Engineering, Southeast University, Nanjing, China wkang@seu.edu.cn

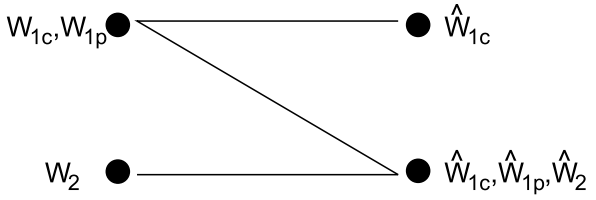

Fig. 3. Possibility 2: The Z channel with degraded message sets.

with the problem of distributed encoding in general and brings us a step closer to eventually solving the capacity of networks containing distributed encoding, like the interference channel, the Z channel, secure communications over the multiple access channel [14], [15] and many others.

II. System Model

Consider a Z channel with degraded message sets, see Figure 3, characterized by $p(y_1|x_1)$ and $p(y_2|x_1,x_2)$ with input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$ and output alphabets $\mathcal{Y}_1$, $\mathcal{Y}_2$. Set

$$V_1(a|b) = \Pr[Y_1 = a|X_1 = b],$$
$$V_2(c|b,d) = \Pr[Y_2 = c|X_1 = b, X_2 = d].$$

There are three independent messages W_{1c}, W_{1p} and W_2 which are uniform on sets $\{1,2,\cdots,M_{1c}\}$, $\{1,2,\cdots,M_{1p}\}$ and $\{1,2,\cdots,M_2\}$, respectively. W_{1c} and W_{1p} is known at Transmitter 1 and W_2 is known at Transmitter 2. W_{1c} is to be decoded at Receiver 1 and W_{1c}, W_{1p}, W_2 are all to be decoded at Receiver 2. An $(M_{1c}, M_{1p}, M_2, n, \epsilon_n)$ code for this channel consists of a sequence of two encoding functions:

$$f_1^n : \{1,2,\cdots,M_{1c}\} \times \{1,2,\cdots,M_{1p}\} \to \mathcal{X}_1^n,$$
$$f_2^n : \{1,2,\cdots,M_2\} \to \mathcal{X}_2^n,$$

and two decoding functions:

$$g_1^n : \mathcal{Y}_1^n \to \{1,2,\cdots,M_{1c}\},$$
$$g_2^n : \mathcal{Y}_2^n \to \{1,2,\cdots,M_{1c}\} \times \{1,2,\cdots,M_{1p}\} \times \{1,2,\cdots,M_2\},$$

with probability of error ϵ_n defined as

$$\frac{1}{M_{1c}M_{1p}M_2} \sum_{w_{1c},w_{1p},w_2} \Pr[g_1^n(Y_1^n) \neq w_{1c}, g_2^n(Y_2^n) \neq (w_{1c},w_{1p},w_2)|W_{1c} = w_{1c}, W_{1p} = w_{1p}, W_2 = w_2].$$

A rate triplet (R_{1c}, R_{1p}, R_2) is said to be achievable if there exists a sequence of $\left(2^{nR_{1c}}, 2^{nR_{1p}}, 2^{nR_2}, n, \epsilon_n\right)$ codes such that $\epsilon_n \to 0$ as $n \to \infty$. The capacity region of the Z channel with degraded message sets is the closure of the set of all achievable rate triplets.

The class of Z channels that we focus on in this paper satisfy the following three conditions. All conditions are placed on the multiple access link between the two transmitters and Receiver 2, i.e., $p(y_2|x_1,x_2)$.

Condition 1: The multiple access link of the Z channel, i.e., $p(y_2|x_1,x_2)$ has the structure shown in Figure 4. More specifically, there exists a random variable T taking values in $\mathcal{T}$ and a deterministic function f such that $p(y_2|x_1,x_2)$ can be expressed as

$$p(y_2|x_1,x_2) = \sum_{t\in\mathcal{T}} p(y_2|x_2,t)p(t|x_1)$$

where $p(y_2|x_2,t)$ is equal to 1 when $y_2 = f(x_2,t)$ and 0 otherwise.

Condition 2: The function $Y_2 = f(X_2,T)$ satisfies that there exists a function h such that $T = h(Y_2,X_2)$.

Condition 3: For any $n = 1,2,\cdots$, the optimal distribution that solves the following optimization problem

$$\max_{p(x_2^n)} \quad H(Y_2^n) \tag{1}$$

is $p^*(x_2^n) = \prod_{i=1}^n p^*(x_{2i})$, i.e., the optimizing distribution is an independent and identically distributed (i.i.d.) distribution according to a single-letter distribution $p^*(x_2)$, irrespective of the distribution $p(t^n)$.

Condition 2 is similar to the conditions on f_1 and f_2 in [16]. From Conditions 1 and 2, we have the following property for any positive integer m and any $p(u,x_1^m)p(x_2^m)p(t^m|x_1^m)$ $p(y_2^m|t^m,x_2^m)$:

$$\begin{aligned} H(Y_2^m|U,X_2^m) &= H(Y_2^m,X_2^m|U,X_2^m) \\ &= H(Y_2^m,X_2^m,T^m|U,X_2^m) \qquad (2) \\ &= H(T^m|U,X_2^m) + H(Y_2^m|U,X_2^m,T^m) \\ &= H(T^m|U) \qquad (3) \end{aligned}$$

where (2) follows because the Z channel satisfies Condition 2, i.e., once we know (X_2,Y_2), we know T, and (3) follows because the Z channel satisfies Condition 1, i.e., the structure of Figure 4.

The main result of this paper is the following characterization of the capacity region of the Z channel with degraded message sets when the multiple access link $p(y_2|x_1,x_2)$ satisfies Conditions 1, 2 and 3.

Theorem 1: For the Z channel with degraded message sets, if the multiple access link $p(y_2|x_1,x_2)$ satisfies Conditions 1, 2 and 3, then the capacity region consists of rate triplets (R_{1c}, R_{1p}, R_2) that satisfies

$$\begin{aligned} R_{1c} &\le I(U;Y_1), \\ R_2 &\le H(Y_2|X_1) - H(T|X_1), \\ R_{1c}+R_{1p} &\le I(U;Y_1) + I(X_1;T|U), \\ R_{1c}+R_{1p} &\le I(X_1;T), \\ R_{1c}+R_{1p}+R_2 &\le H(Y_2|U) - H(T|X_1) + I(U;Y_1), \\ R_{1c}+R_{1p}+R_2 &\le H(Y_2) - H(T|X_1), \end{aligned}$$

for some $p(u,x_1)$, with $|\mathcal{U}| \le |\mathcal{X}_1|+2$, where $|\mathcal{U}|$ is the cardinality of the random variable U and the mutual informations are evaluated using the distribution $p(u,x_1,x_2,t,y_1,y_2) = p(u,x_1)p^*(x_2)p(t|x_1)p(y_1|x_1)p(y_2|t,x_2)$.

The achievability and converse parts of Theorem 1 is proved in Sections IV and V, respectively. The cardinality of the auxiliary random variable U follows from the support lemma [12, Lemma 3.4].

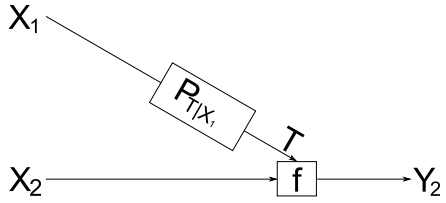

Fig. 4. The multiple access link $p(y_2|x_1, x_2)$.

III. Examples

In this section, we provide two examples of the multiple access link $p(y_2|x_1, x_2)$ that satisfies Conditions 1, 2 and 3.

A. The Modulo-sum Channel

For the modulo-sum channel [11], $\mathcal{X}_2 = \mathcal{T} = \mathcal{Y}_2 = \{0, 1, 2, \cdots, q-1\}$, and $p(t|x_1)$ and $\mathcal{X}_1$ are both arbitrary. Y_2 can be written as $Y_2 = X_2 \oplus T$, where $\oplus$ represents the modulo-sum. From the description of the channel, the modulo-sum channel clearly satisfies Condition 1. Since we have $T = Y_2 \oplus X_2$, the modulo-sum channel satisfies Condition 2. As for Condition 3, the maximum value of $H(Y_2^n)$ over all distributions $p(x_2^n)p(t^n)$ is $n \log q$, and this is achieved when $p(x_2^n)$ is the i.i.d distribution according to the uniform distribution on $\{0, 1, 2, \cdots, q-1\}$ irrespective of $p(t^n)$. Hence, the modulo-sum channel satisfies Condition 3 as well.

B. The Erasure Channel

For the erasure channel, $\mathcal{X}_2 = \{0, 1\}$, $\mathcal{T} = \{0, e\}$, $\mathcal{Y}_2 = \{0, e, 1\}$, and $p(t|x_1)$ and $\mathcal{X}_1$ are both arbitrary. Y_2 can be written as $Y_2 = f(X_2, T)$ where the mapping f is

$$f(0,0) = 0, \quad f(0,e) = e, \quad f(1,0) = 1, \quad f(1,e) = e.$$

As can be seen, the channel between X_2 to Y_2 is an erasure channel where $\Pr[T = e]$ models the probability of erasure.

It is easy to see that the erasure channel satisfies Conditons 1 and 2. We now argue that the uniform distribution on $\mathcal{X}_2^n = \{0, 1\}^n$ achieves the maximization in (1), which means that the erasure channel also satisfies Condition 3. We have

$$\begin{aligned} H(Y_2^n) &= H(Y_2^n, T^n) \quad (4)\\ &= H(T^n) + H(Y_2^n|T^n) \\ &= H(T^n) + \sum_{t^n \in \{0,e\}^n} H(Y_2^n|T^n = t^n)\Pr[T^n = t^n], \quad (5) \end{aligned}$$

where (4) follows because the function f is such that once we know $f(\cdot, \omega)$, we know ω. we will next show that the uniform distribution on $\mathcal{X}_2^n = \{0, 1\}^n$ maximizes $H(Y_2^n|T^n = t^n)$ for each $t^n \in \{0, e\}^n$. Let t^n contain k number of e's, $0 \le k \le n$. Then, given $T^n = t^n$, Y_2^n can take at most 2^{n-k} many values, thus,

$$H(Y_2^n|T^n = t^n) \le n - k \quad (6)$$

with equality when X_2^n is the uniform distribution on $\mathcal{X}_2^n = \{0, 1\}^n$. Hence, when X_2^n takes the uniform distribution, $H(Y_2^n|T^n = t^n)$ is maximized for each $t^n \in \{0, e\}^n$ and therefore, according to (5), $H(Y_2^n)$ is maximized. Thus, the erasure channel satisfies Conditions 1, 2 and 3 in this paper.

Remark: For Gaussian additive channels, $Y_2 = X_1 + X_2 + Z$, where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \mathbb{R}$ and Z is a zero-mean Gaussian random variable with unit variance, it satisfies Conditions 1 and 2, but not Condition 3. To see this, even for $n = 1$, the distribution that maximizes (1), subject to power constraint P_2, depends on the distribution of X_1 and the Gaussian distribution with zero mean and P_2 variance can not always be the optimal distribution.

IV. Achievability

In this section, we prove the achievability part of Theorem 1, i.e., we show that the region described in Theorem 1 is achievable.

We first provide an achievability result for the general Z channel with degraded message sets. The achievability scheme is as follows: superposition encoding is used at Transmitter 1. It encodes W_{1c} and part of W_{1p} into the inner codeword and the remaining part of W_{1p} into the outer codeword. Transmitter 2 uses an i.i.d. codebook for its message W_2. Receiver 1 decodes the inner codeword using joint typicality decoding while treating everything else as noise, and Receiver 2 decodes the inner codeword, the outer codeword and the codeword of Transmitter 2 using joint typicality decoding.

Theorem 2: For the Z channel with degraded message sets, rate triplets (R_{1c}, R_{1p}, R_2) that satisfies

$$\begin{aligned} R_{1c} &\le I(U; Y_1|Q), \\ R_2 &\le I(X_2; Y_2|X_1, Q), \\ R_{1c} + R_{1p} &\le I(U; Y_1|Q) + I(X_1; Y_2|X_2, U, Q), \\ R_{1c} + R_{1p} &\le I(X_1; Y_2|X_2, Q), \\ R_{1c} + R_{1p} + R_2 &\le I(X_1, X_2; Y_2|U, Q) + I(U; Y_1|Q), \\ R_{1c} + R_{1p} + R_2 &\le I(X_1, X_2; Y_2|Q), \end{aligned}$$

for some $p(q)p(u, x_1|q)p(x_2|q)$ are achievable, where the mutual informations are evaluated using the distribution $p(q, u, x_1, x_2, y_1, y_2) = p(q)p(u, x_1|q)p(x_2|q)V_1(y_1|x_1)$ $V_2(y_2|x_1, x_2)$.

Proof: The details of the proof are provided in the appendix. ∎

If we set $p(x_2|q)$ in the achievable region in Theorem 2 to be a specific distribution, we obtain yet another achievable region, potentially smaller. For the Z channel with degraded message sets that satisfies Conditions 3, set $p(x_2|q)$ to be $p^*(x_2)$ and omit time-sharing, using the fact that the channel satisfies (3), we have shown that the region described in Theorem 1 is achievable. The achievability part of the proof of Theorem 1 is thus complete.

V. Converse

In this section, we provide the converse part of the proof of Theorem 1, i.e., we show that if the Z channel with degraded message sets satisfies Conditions 1, 2 and 3, then all achievable rate triplets (R_{1c}, R_{1p}, R_2) are contained in the region described in Theorem 1.

For any rate triplet (R_{1c}, R_{1p}, R_2) that is achievable, there exist two sequences of codebooks 1 and 2, denoted by $\mathcal{C}_1^n$ and $\mathcal{C}_2^n$, of rates $R_{1c}+R_{1p}$ and R_2, and probability of error less than ϵ_n, where $\epsilon_n \to 0$ as $n \to \infty$. Let X_1^n and X_2^n be uniformly distributed on codebooks 1 and 2, respectively. Let Y_1^n be connected via V_1^n to X_1^n, T^n be connected via $p(t|x_1)$ to X_1^n, Y_2^n be connected via $p(y_2|t,x_2)$ to T^n and X_2^n. Similar to the converse of the multiple access channel [17], we have

$$\begin{aligned} nR_2 &= H(W_2|W_{1c}, W_{1p}) \\ &\le I(X_2^n; Y_2^n|X_1^n) + n\epsilon_n \qquad (7) \\ &\le \sum_{i=1}^n I(X_{2i}; Y_{2i}|X_{1i}) + n\epsilon_n, \\ &\le \sum_{i=1}^n H(Y_{2i}|X_{1i}) - H(T_i|X_{1i}), \qquad (8) \end{aligned}$$

and

$$\begin{aligned} nR_{1c} + nR_{1p} &= H(W_{1c}, W_{1p}|W_2) \\ &\le I(X_1^n; Y_2^n|X_2^n) + n\epsilon_n \qquad (9) \\ &\le \sum_{i=1}^n I(X_{1i}; Y_{2i}|X_{2i}) + n\epsilon_n, \\ &\le \sum_{i=1}^n I(X_{1i}; T_i) + n\epsilon_n, \qquad (10) \end{aligned}$$

and

$$\begin{aligned} nR_{1c} + nR_{1p} + nR_2 &= H(W_{1c}, W_{1p}, W_2) \\ &\le I(X_1^n, X_2^n; Y_2^n) + n\epsilon_n \qquad (11) \\ &\le \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_{2i}) + n\epsilon_n, \\ &\le \sum_{i=1}^n H(Y_{2i}) - H(T_i|X_{1i}) + n\epsilon_n, \qquad (12) \end{aligned}$$

where (7), (9) and (11) follow from Fano's inequality, and (8), (10), and (12) follows from fact that the channel satisfies (3). As for the other constraints on the achievable rate triplets, we have

$$\begin{aligned} nR_{1c} &= H(W_{1c}) \\ &\le I(W_{1c}; Y_1^n) + n\epsilon_n \qquad (13) \\ &= H(Y_1^n) - H(Y_1^n|W_{1c}) + n\epsilon_n \\ &\le \left[\sum_{i=1}^n H(Y_{1i})\right] - H(Y_1^n|W_{1c}) + n\epsilon_n, \qquad (14) \end{aligned}$$

where (13) follows from Fano's inequality and (14) follows from the chain rule and conditioning reduces entropy. We also have

$$\begin{aligned} nR_{1p} &= H(W_{1p}|W_{1c}, W_2) \\ &\le I(W_{1p}; Y_2^n|W_2, W_{1c}) + n\epsilon_n \qquad (15) \\ &= H(Y_2^n|W_2, W_{1c}) - H(Y_2^n|W_{1p}, W_{1c}, W_2) + n\epsilon_n \\ &= H(Y_2^n|X_2^n, W_{1c}) - H(Y_2^n|X_1^n, X_2^n) + n\epsilon_n \qquad (16) \\ &= H(Y_2^n|X_2^n, W_{1c}) - \sum_{i=1}^n H(Y_{2i}|X_{1i}, X_{2i}) + n\epsilon_n, \qquad (17) \\ &= H(T^n|W_{1c}) - \sum_{i=1}^n H(T_i|X_{1i}) + n\epsilon_n, \qquad (18) \end{aligned}$$

where (15) follows from Fano's inequality, (16) follows from the fact that without loss of generality, we may consider deterministic encoders f_1^n and f_2^n, (17) follows from the memoryless nature of the channel $p(y_2|x_1,x_2)$, and (18) follows because the channel satisfies (3). We further have

$$\begin{aligned} nR_{1p} + nR_2 &= H(W_{1p}, W_2|W_{1c}) \\ &\le I(W_{1p}, W_2; Y_2^n|W_{1c}) + n\epsilon_n \qquad (19) \\ &= H(Y_2^n|W_{1c}) - H(Y_2^n|W_{1p}, W_{1c}, W_2) + n\epsilon_n \\ &= H(Y_2^n|W_{1c}) - H(Y_2^n|X_1^n, X_2^n) + n\epsilon_n \\ &= H(Y_2^n|W_{1c}) - \sum_{i=1}^n H(Y_{2i}|X_{1i}, X_{2i}) + n\epsilon_n \\ &= H(Y_2^n|W_{1c}) - \sum_{i=1}^n H(T_i|X_{1i}) + n\epsilon_n, \qquad (20) \end{aligned}$$

where (19) follows from Fano's inequality, and (20) follows because the channel satisfies (3).

Next, we define imaginary channels and random variables. The purpose of doing this is to utilize Condition 3 and define auxiliary random variables during single-letterization that are independent to X_2^n. Define $\tilde{X}_2^n$ to be i.i.d. according to the distribution $p^*(x_2)$ and is independent to everything else. Further define $\tilde{Y}_2^n$ to be the output of T^n and $\tilde{X}_2^n$ into the memoryless channel $p(y_2|t,x_2)$. Due to condition 3, we have

$$H(Y_2^n|W_{1c}) \le H(\tilde{Y}_2^n|W_{1c}), \qquad (21)$$

because for each $W_{1c} = w_{1c}$, $w_{1c} \in \{1, 2, \cdots, 2^{nR_{1c}}\}$, $H(Y_2^n|W_{1c}=w_{1c}) \le H(\tilde{Y}_2^n|W_{1c}=w_{1c})$ since for $p(x_1^n|W_{1c}=w_{1c})$, the maximizing $p(x_2^n)$ is the i.i.d. distribution according to $p^*(x_2)$. Note that $\tilde{Y}_2^n$ is independent to X_2^n.

From (20) and (21), we have

$$nR_{1p} + nR_2 \le H(\tilde{Y}_2^n|W_{1c}) - \sum_{i=1}^n H(T_i|X_{1i}) + n\epsilon_n. \qquad (22)$$

Next, we proceed with the single-letterization of n-letter

entropies using [12, page 314, eqn (3.34)]:

$$H(\tilde{Y}_2^n|W_{1c}) - H(Y_1^n|W_{1c}) = \sum_{i=1}^{n} \Big[H(\tilde{Y}_{2i}|V_i) - H(Y_{1i}|V_i) \Big], \quad (23)$$

$$H(T^n|W_{1c}) - H(Y_1^n|W_{1c}) = \sum_{i=1}^{n} \Big[H(T_i|U_i) - H(Y_{1i}|U_i) \Big], \quad (24)$$

where auxiliary random variables V_i and U_i, $i = 1, 2, \cdots, n$ are defined as

$$V_i = (Y_1^{i-1}, \tilde{Y}_{2(i+1)}^n, W_{1c}), \quad U_i = (Y_1^{i-1}, T_{(i+1)}^n, W_{1c}). \quad (25)$$

Note that U_i depends only on X_1^n and W_{1c}, while V_i depends only on X_1^n, W_{1c} and $\tilde{X}_{2(i+1)}^n$. Since $\tilde{X}_2^n$ is i.i.d. and independent to everything else, $\tilde{X}_{2i}$ is independent to (U_i, V_i). Based on the definition of $\tilde{Y}_2^n$, we have the following Markov chains which will be used later:

$$(Y_{1i}, \tilde{Y}_{2i}, T_i) \to U_i \to V_i \quad (26)$$

$$(U_i, V_i) \to T_i \to \tilde{Y}_{2i} \quad (27)$$

From (14) we have

$$\begin{aligned} nR_{1c} &\le \left[\sum_{i=1}^{n} H(Y_{1i}) \right] - H(Y_1^n|W_{1c}) + n\epsilon_n \\ &= \sum_{i=1}^{n} H(Y_{1i}) - \sum_{i=1}^{n} H(Y_{1i}|Y_1^{i-1}, W_{1c}) + n\epsilon_n \\ &\le \sum_{i=1}^{n} H(Y_{1i}) - \sum_{i=1}^{n} H(Y_{1i}|Y_1^{i-1}, W_{1c}, \tilde{Y}_{2(i+1)}^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} I(V_i; Y_{1i}) + n\epsilon_n, \end{aligned} \quad (28)$$

where (28) follows from the definition of V_i in (25). From (14) and (18) we have

$$\begin{aligned} &nR_{1c} + nR_{1p} \\ &\le \left[\sum_{i=1}^{n} H(Y_{1i}) \right] - H(Y_1^n|W_{1c}) + H(T^n|W_{1c}) \\ &\quad - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n \\ &= \sum_{i=1}^{n} H(Y_{1i}) + \sum_{i=1}^{n} H(T_i|U_i) - \sum_{i=1}^{n} H(Y_{1i}|U_i) \\ &\quad - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n \quad (29) \\ &= \sum_{i=1}^{n} I(U_i; Y_{1i}) + \sum_{i=1}^{n} H(T_i|U_i) - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n, \end{aligned} \quad (30)$$

where (29) follows from (24). From (14) and (22) we have

$$\begin{aligned} &nR_{1c} + nR_{1p} + nR_2 \\ &\le \left[\sum_{i=1}^{n} H(Y_{1i}) \right] - H(Y_1^n|W_{1c}) + H(\tilde{Y}_2^n|W_{1c}) \\ &\quad - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n \\ &= \sum_{i=1}^{n} H(Y_{1i}) - \sum_{i=1}^{n} H(Y_{1i}|V_i) + \sum_{i=1}^{n} H(\tilde{Y}_{2i}|V_i) \\ &\quad - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n \quad (31) \\ &= \sum_{i=1}^{n} I(V_i; Y_{1i}) + \sum_{i=1}^{n} H(\tilde{Y}_{2i}|V_i) - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n, \end{aligned} \quad (32)$$

where (31) follows from (23).

We have defined two sequences of auxiliary random variables in the converse, namely, U_i and V_i, while in the achievability results in Theorem 2, there is just one auxiliary random variable. So next, using the technique from [13], we would like to replace U_i and V_i with one auxiliary random variable S_i, $i = 1, 2, \cdots, n$. In other words, we would like to show that for every $i = 1, 2, \cdots, n$, there exists a random variable S_i such that

$$I(V_i; Y_{1i}) \le I(S_i; Y_{1i}), \quad (33)$$

$$I(U_i; Y_{1i}) + H(T_i|U_i) \le I(S_i; Y_{1i}) + H(T_i|S_i), \quad (34)$$

$$I(V_i; Y_{1i}) + H(\tilde{Y}_{2i}|V_i) \le I(S_i; Y_{1i}) + H(\tilde{Y}_{2i}|S_i). \quad (35)$$

Once we have such S_i, we may further upper bound (28), (30) and (32) and get rid of auxiliary random variables U_i and V_i. If U_i and V_i is such that

$$I(\tilde{Y}_{2i}; U_i|V_i) \le I(Y_{1i}; U_i|V_i), \quad (36)$$

choose S_i to be U_i, then (34) is naturally satisfied while (33) is satisfied because of (26). (35) is also satisfied because

$$\begin{aligned} &I(V_i; Y_{1i}) + H(\tilde{Y}_{2i}|V_i) - I(U_i; Y_{1i}) - H(\tilde{Y}_{2i}|U_i) \\ &= I(\tilde{Y}_{2i}; U_i|V_i) - I(Y_{1i}; U_i|V_i) \\ &\le 0, \end{aligned}$$

since U_i and V_i satisfy (36). If U_i and V_i is such that (36) is not true, i.e.,

$$I(\tilde{Y}_{2i}; U_i|V_i) > I(Y_{1i}; U_i|V_i), \quad (37)$$

choose S_i to be V_i, then (33) and (35) are naturally satisfied while (34) is satisfied because

$$\begin{aligned} &I(U_i; Y_{1i}) + H(T_i|U_i) - I(V_i; Y_{1i}) - H(T_i|V_i) \\ &= I(U_i; Y_{1i}|V_i) - I(U_i; T_i|V_i) \\ &\le I(U_i; Y_{1i}|V_i) - I(U_i; \tilde{Y}_{2i}|V_i) \quad (38) \\ &< 0, \quad (39) \end{aligned}$$

where (38) follows because of (27), and (39) follows because of (37). Thus, from (28), (30), (32) combined with (33)-(35), we have

$$nR_{1c} \le \sum_{i=1}^{n} I(S_i;Y_{1i}) + n\epsilon_n, \tag{40}$$

$$nR_{1c} + nR_{1p} \le \sum_{i=1}^{n} I(S_i;Y_{1i}) + \sum_{i=1}^{n} H(T_i|S_i) - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n, \tag{41}$$

$$nR_{1c} + nR_{1p} + nR_2 \le \sum_{i=1}^{n} I(S_i;Y_{1i}) + \sum_{i=1}^{n} H(\tilde{Y}_{2i}|S_i) - \sum_{i=1}^{n} H(T_i|X_{1i}) + 2n\epsilon_n. \tag{42}$$

Further define Q to be a random variable uniform on $\{1, 2, \cdots, n\}$ and is independent to everything else. Define

$$S = S_Q, X_1 = X_{1Q}, X_2 = X_{2Q}, Y_1 = Y_{1Q}, Y_2 = Y_{2Q},$$
$$T = T_Q, \tilde{Y}_2 = \tilde{Y}_{2Q}, \tilde{X}_2 = \tilde{X}_{2Q}.$$

The auxiliary random variables thus defined satisfies

$$p(q,s,x_1,x_2,\tilde{x}_2,y_1,y_2,\tilde{y}_2,t) = p(q,s,x_1)p(x_2|q)p^*(\tilde{x}_2) p(y_1|x_1)p(t|x_1)p(\tilde{y}_2|t,\tilde{x}_2)p(y_2|t,x_2). \tag{43}$$

In (43), $(S, X_1) \to Q \to X_2$ follows because the definition of S_i only depends on U_i and V_i, which in turn only depends on W_{1c}, Y_1^{i-1}, T_{i+1}^n and $\tilde{Y}_{2(i+1)}^n$, which in turn only depends on W_{1c}, X_1^n and $\tilde{X}_{2(i+1)}^n$, all of which are independent to X_2^n.

Hence, from (8), (10), (12) and (40)-(42), letting $n \to \infty$, we have shown that the achievable rate triplets (R_{1c}, R_{1p}, R_2) have to satisfy

$$R_{1c} \le I(S;Y_1|Q), \tag{44}$$

$$R_2 \le H(Y_2|X_1,Q) - H(T|X_1), \tag{45}$$

$$R_{1c} + R_{1p} \le I(S;Y_1|Q) + H(T|S,Q) - H(T|X_1), \tag{46}$$

$$R_{1c} + R_{1p} \le I(X_1;T|Q), \tag{47}$$

$$R_{1c} + R_{1p} + R_2 \le H(\tilde{Y}_2|S,Q) - H(T|X_1) + I(S;Y_1|Q), \tag{48}$$

$$R_{1c} + R_{1p} + R_2 \le H(Y_2|Q) - H(T|X_1), \tag{49}$$

for some $p(q,s,x_1)p(x_2|q)$ where the joint distribution of the random variables satisfy (43).

Now, we perform a series of transformations on the region described in (44)-(49) so that it resembles the region in Theorem 1. First, we would like to remove the variable $\tilde{Y}_2$. Notice that in the region (44)-(49), only the term $H(Y_2|X_1,Q)$ in (45) and the term $H(Y_2|Q)$ in (49) depend on the distribution of $p(x_2|q)$. Due to Condition 3, both of these terms are maximized when using the $p^*(x_2)$ distribution. Thus, using any other $p(x_2|q)$ distribution offers no enlargement of the region. Thus, we may restrict ourselves to using $p^*(x_2)$ and the above region is unchanged if we replace the distribution in (43) with the distribution $p(q,s,x_1,x_2,t,y_1,y_2,\tilde{y}_2,\tilde{x}_2) = p(q)p^*(\tilde{x}_2)p(s,x_1|q)p^*(x_2)p(t|x_1)p(y_1|x_1)p(y_2|t,x_2)$ $p(\tilde{y}_2|t,\tilde{x}_2)$.

Since now X_2 and $\tilde{X}_2$ take the same distribution, we have $H(Y_2|S,Q) = H(\tilde{Y}_2|S,Q)$ for the same $p(q,s,x_1)$. Hence, we may write the region by replacing $H(\tilde{Y}_2|S,Q)$ by $H(Y_2|S,Q)$ in (48) and get

$$R_{1c} \le I(S;Y_1|Q), \tag{50}$$

$$R_2 \le H(Y_2|X_1,Q) - H(T|X_1), \tag{51}$$

$$R_{1c} + R_{1p} \le I(S;Y_1|Q) + I(X_1;T|S,Q), \tag{52}$$

$$R_{1c} + R_{1p} \le I(X_1;T|Q), \tag{53}$$

$$R_{1c} + R_{1p} + R_2 \le H(Y_2|S,Q) - H(T|X_1) + I(S;Y_1|Q), \tag{54}$$

$$R_{1c} + R_{1p} + R_2 \le H(Y_2|Q) - H(T|X_1), \tag{55}$$

for some $p(q,s,x_1)$ where the distributions are calculated according to $p(q,s,x_1,x_2,t,y_1,y_2) = p(q,s,x_1)p^*(x_2)p(t|x_1)p(y_1|x_1)p(y_2|t,x_2)$.

Next, since $p(x_2|q)$ is specified to be $p^*(x_2)$ irrespective of q, we further upper bound the rates in (50)-(55) to remove the time-sharing random variable Q: we upper bound $I(S;Y_1|Q)$ by $I(Q,S;Y_1)$ in (50), (52) and (54), $H(Y_2|X_1,Q)$ by $H(Y_2|X_1)$ in (51), $H(T|Q)$ by $H(T)$ in (53), and $H(Y_2|Q)$ by $H(Y_2)$ in (55). Further define auxiliary random variable $U = (Q,S)$, we obtain

$$R_{1c} \le I(U;Y_1), \tag{56}$$

$$R_2 \le H(Y_2|X_1) - H(T|X_1), \tag{57}$$

$$R_{1c} + R_{1p} \le I(U;Y_1) + I(X_1;T|U), \tag{58}$$

$$R_{1c} + R_{1p} \le H(T) - H(T|X_1), \tag{59}$$

$$R_{1c} + R_{1p} + R_2 \le H(Y_2|U) - H(T|X_1) + I(U;Y_1), \tag{60}$$

$$R_{1c} + R_{1p} + R_2 \le H(Y_2) - H(T|X_1), \tag{61}$$

for some $p(u,x_1)$ where the distributions are calculated according to $p(u,x_1,x_2,t,y_1,y_2) = p(u,x_1)p^*(x_2)p(t|x_1)p(y_1|x_1)p(y_2|t,x_2)$. This shows that the achievable rate triplets (R_{1c}, R_{1p}, R_2) have to lie within the region described in Theorem 1, which completes the converse part of the proof of Theorem 1.

VI. Conclusions

When the multiple access link between the two transmitters and Receiver 2 satisfies certain conditions, we have characterized the capacity region of the Z channel with degraded message sets despite the presence of distributed encoding. The conditions and techniques presented in this paper may be useful in obtaining capacity results for other networks with distributed encoding.

Appendix

Proof of Theorm 2

Fix $p(q)p(u,x_1|q)p(x_2|q)$ and $\gamma \in [0, R_{1p}]$.

Random codebook generation: First randomly generate a time-sharing sequence q^n in an i.i.d. fashion according to the distribution $p(q)$. The codebook at Transmitter 1 is generated as follows: conditioned on the q^n sequence, randomly generate $2^{n(R_{1c}+\gamma)}$ codewords in a conditional i.i.d. fashion according to the distribution $p(u|q)$. These constitute the inner codebook and each inner codeword is denoted as $u^n(i)$, $i = 1, 2, \cdots 2^{n(R_{1c}+\gamma)}$. Conditioned on inner codeword i, $i = 1, 2, \cdots 2^{n(R_{1c}+\gamma)}$, and q^n, randomly generate $2^{n(R_{1p}-\gamma)}$ codewords in an conditional i.i.d. fashion according to the distribution $p(x_1|u, q)$. These constitute the i-th outer codebook and each outer codeword is denoted as $x_1^n(i, j)$, $i = 1, 2, \cdots 2^{n(R_{1c}+\gamma)}$, $j = 1, 2, \cdots, 2^{n(R_{1p}-\gamma)}$. The codebook at Transmitter 2 is generated as follows: conditioned on the q^n sequence, randomly generate 2^{nR_2} many codewords in a conditional i.i.d. fashion according to $p(x_2|q)$. Each codeword is denoted as $x_2^n(m)$, $m = 1, 2, \cdots, 2^{nR_2}$. The codebooks, as well as q^n, is revealed to the encoders and decoders.

Encoding: Transmitter 1 splits its message W_{1p} into two parts W_{1pa} and W_{1pb} with rates γ and $R_{1p}-\gamma$, respectively. Suppose $W_{1c} = w_{1c}$, $W_{1pa} = w_{1pa}$, $W_{1pb} = w_{1pb}$ and $W_2 = w_2$, Transmitter 1 transmits $x_1^n(w_{1pa} \cdot 2^{nR_{1c}} + w_{1c}, w_{1pb})$, and Transmitter 2 transmits $x_2^n(w_2)$.

Decoding: Receiver 1 finds the unique index $\bar{i}$ which satisfies that $(q^n, u^n(\bar{i}), Y_1^n)$ is in the joint typical set associated with the distribution $p(q, u, y_1)$. If no such index exists or if there are more than one, it outputs an error. Receiver 2 finds the unique triplet of indices $(\bar{i}, \bar{j}, \bar{m})$ which satisfies that $(q^n, u^n(\bar{i}), x_1^n(\bar{i}, \bar{j}), x_2^n(\bar{m}), Y_2^n)$ is in the joint typical set associated with the distribution $p(q, u, x_1, x_2, y_2)$. If no such triplet exists or if there are more than one, it outputs an error.

Probability of error calculation: Due the symmetry of the codebook design, the average probability of error is the same as the probability of error if we assume that $x_1^n(1, 1)$ and $x_2^n(1)$ are transmitted. Error occurs only if one of the following events happen:

1) E_{11}:$(q^n, u^n(1), Y_1^n)$ is not jointly typical. Due to the asymptotic equipartition property (AEP) [17], the probability of this event happening goes to zero as n goes to infinity.
2) E_{12}: There exists some $i \neq 1$ such that $(q^n, u^n(i), Y_1^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_{1c}+\gamma-I(U;Y_1|Q)+\epsilon)}$.
3) E_{21}: $(q^n, u^n(1), x_1^n(1, 1), x_2^n(1), Y_2^n)$ is not jointly typical. Due to AEP, the probability of this event happening goes to zero as n goes to infinity.
4) E_{22}: There exists some $m \neq 1$ such that $(q^n, u^n(1), x_1^n(1, 1), x_2^n(m), Y_2^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_2-I(X_2;Y_2|X_1,Q)+\epsilon)}$.
5) E_{23}: There exists some $j \neq 1$ such that $(q^n, u^n(1), x_1^n(1, j), x_2^n(1), Y_2^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_{1p}-\gamma-I(X_1;Y_2|X_2,U,Q)+\epsilon)}$.
6) E_{24}: There exists some $j \neq 1, m \neq 1$ such that $(q^n, u^n(1), x_1^n(1, j), x_2^n(m), Y_2^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_2+R_{1p}-\gamma-I(X_1,X_2;Y_2|U,Q)+\epsilon)}$.
7) E_{25}: There exists some $i \neq 1, j \neq 1$ such that $(q^n, u^n(i), x_1^n(i, j), x_2^n(1), Y_2^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_{1c}+R_{1p}-I(X_1;Y_2|X_2,Q)+\epsilon)}$.
8) E_{26}: There exists some $i \neq 1, j \neq 1, m \neq 1$ such that $(q^n, u^n(i), x_1^n(i, j), x_2^n(m), Y_2^n)$ is jointly typical. The probability of this event happening is upper bounded by $2^{n(R_{1c}+R_{1p}+R_2-I(X_1,X_2;Y_2|Q)+\epsilon)}$.

Using the union bound and Fourier-Motzkin elimination, we obtain the desired results.

REFERENCES

[1] S. Vishwanath, N. Jindal, and A. Goldsmith. The "Z" channel. In *IEEE Globecom*, San Francisco, CA, December 2003.

[2] N. Liu and S. Ulukus. On the capacity of the Gaussian Z-channel. In *IEEE Global Communications Conference*, Dallas, TX, November 2004.

[3] H. F. Chong, M. Motani, and H.K. Garg. Capacity theorems for the Gaussian zigzag channel. In *IEEE International Symposium on Information Theory*, Seattle, WA, July 2006.

[4] H. F. Chong, M. Motani, and H.K. Garg. Capacity theorems for the "Z" channel. *IEEE Trans. on Information Theory*, 53(4):1348–1365, April 2007.

[5] S. A. Jafar V. R. Cadambe and S. Vishwanath. The capacity region of a class of deterministic Z channels. In *IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009.

[6] S. Salehkalaibar and M. R. Aref. On the capacity region of the degraded Z channel. In *IEEE Information Theory Workshop*, Dublin, Ireland, September 2010.

[7] W. Liu, N. Liu, and Z. Pan. On the capacity region of the semi-deterministic Z channel. In *Proceedings of the International Conference on Wireless Communications and Signal Processing*, Nanjing, China, November 2011.

[8] T. J Oechtering and M. Skoglund. Capacity bounds for the Z channel. In *Proc. IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, October 2011.

[9] J. Korner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. on Information Theory*, 23(1):60–64, Jan. 1977.

[10] N. Liu, D. Gunduz, A. Goldsmith, and H. V. Poor. Interference channels with correlated receiver side information. *IEEE Trans. Information Theory*, 56(12):5984–5998, December 2010.

[11] N. Liu and A. Goldsmith. Capacity regions and bounds for a class of Z-interference channels. *IEEE Trans. on Information Theory*, 55(11):4986–4994, November 2009.

[12] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[13] A. Lapidoth and L. Wong. The state-dependent semideterministic broadcast channel. *Available at http://arxiv.org/abs/1111.1144*.

[14] Y. Liang and H. V. Poor. Generalized multiple access channels with confidential messages. *IEEE Trans. on Information Theory*, 54(3):976–1002, March 2008.

[15] E. Tekin and A. Yener. The Gaussian multiple access wire-tap channel. *IEEE Trans. on Information Theory*, 54(12):5747–5755, December 2008.

[16] A. El Gamal and M. Costa. The capacity region of a class of deterministic interference channels. *IEEE Trans. on Information Theory*, 28(2):343–346, March 1982.

[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.