State-Dependent Z Channel

S. Hajizadeh, M. Monemizadeh, E. Bahmani, G. A. Hodtani and M. Joneidi

1 Department of Electrical Engineering, University of Illinois at Chicago, Chicago, IL
2 Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
3 Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Abstract—In this paper we study the “Z” channel with side information non-causally available at the encoders. We use Marton encoding along with Gelfand-Pinsker random binning scheme and Chong-Motani-Garg-El Gamal (CMGE) jointly decoding to find an achievable rate region. We will see that our achievable rate region gives the achievable rate of the multiple access channel with side information and also degraded broadcast channel with side information. We will also derive an inner bound and an outer bound on the capacity region of the state-dependent degraded discrete memoryless Z channel. We will then observe that using Costa dirty paper coding, we can remove the bad effect of the interference from the direction of one transmitter-receiver pair. Also, by assuming the high signal to noise ratio and strong interference regime, and using the lattice strategies, we derive an achievable rate region for the Gaussian degraded “Z” channel with additive interference non-causally available at both of the encoders. Our method is based on lattice transmission scheme, jointly decoding at the first decoder and successive decoding at the second decoder. Using such coding scheme we remove the effect of the interference completely.

Keywords: Z channel; Side information; rate splitting; Dirty paper coding; Lattice strategies

I. INTRODUCTION

The Z channel is a two-transmitter two-receiver model shown in Fig. 1 where the first sender only wishes to send information to the first receiver whereas the second transmitter sends information to both of the receivers. The Z channel was first studied by Viswanath et al. [1] where they introduced the model and found the capacity region of a special class of Z channels and the achievable rate of a special case of the Gaussian Z channel (GZC). In [2], Liu and Ulukus obtained several capacity bounds for a class of GZC. Chong-Motani-Garg (CMGE) [3] studied three different types of degraded Z channel and characterized the capacity region in one type. They also characterized the capacity region of GZC with moderately strong crossover link.

The capacity region of the general Z channel is still an open problem. The best achievable rate region for the discrete memoryless Z channel until today is due to Do et al. [4].

Channels with side information were first studied by Shannon [5] where he characterized the capacity of a point-to-point channel with side information causally available at the transmitters. Gelfand and Pinsker [6] found the capacity of a single-user channel with side information non-causally available at the encoders.

In this paper we study the Z channel with channel state information non-causally available at the encoders that is depicted in Fig. 2. The reason to study this channel model is buttressed by the applications it has in some wireless communication scenarios such as the case where two communication-involved cells are interfering with each other and thus suffer from a common interference modeled by some S non-causally available to two distinct destination base stations as shown in Fig. 3.

As in Fig. 2, the first transmitter sends $m_{11} \in [1; 2^{nR_{11}}]$ to $Y_1$ while the second transmitter first splits its messages, $m_{21} \in [1; 2^{nR_{21}}]$ and $m_{22} \in [1; 2^{nR_{22}}]$, to two independent parts; i.e. $M_{21} = (M_{211}, M_{212})$ and $M_{22} = (M_{221}, M_{222})$ with rates $R_{21} = R_{211} + R_{212}$ and $R_{22} = R_{221} + R_{222}$ respectively, and then encodes its messages to send through the
channel. The channel state information is non-causally available at the transmitters. The messages $M_{2k}^R$ can be decoded by both receivers, while $M_{2k}^B$ is decoded by its respective receiver, $k = 1, 2.$

We propose an achievable rate region using the lattice based coding for the Gaussian degraded “Z” channel with additive interference non-causally available at both of the encoders under high-SNR and strong interference regime. Our method is based on lattice transmission scheme, jointly decoding at the first decoder and successive decoding at the second decoder. Using such coding scheme we remove the effect of the interference completely.

The rest of the paper is as follows. In section II, definitions are provided. In section III, we derive an achievable rate region for the general discrete memoryless Z channel with side information non-causally available at the encoders. In section IV, we derive an inner and an outer bound on the capacity region of degraded discrete memoryless Z channel and we will observe that our outer bound coincides with the inner bound for the communication rates of the second transmitter, i.e. $R_{21}$ and $R_{22}.$ We will also show that using dirty paper coding, we can remove the negative effect of the interference in the direction of one transmitter-receiver pair in the derived inner bound. In section V, we derive an achievable rate region for the Gaussian degraded “Z” channel with additive interference non-causally available at both of the encoders using lattice strategies and show that using lattice strategy we can completely remove the interference. The conclusion is given in section VI.

II. DEFINITIONS

The discrete memoryless “Z” channel with channel state information non-causally available at the transmitter, depicted in Fig. 2, consists of five finite sets $S, X_1, X_2, Y_1, Y_2$ and two marginal probability distributions $p(y_1|y_2, s)$ and $p(y_1|x_2, s).$ The memorylessness nature of the channel imposes the following additional constraint on the channel transition probability

$$p(y_1^n, y_2^n|x_1^n, x_2^n, s^n) = \prod_{i=1}^{n} p(y_{2i}|x_{2i}, s_i)p(y_{1i}|x_{1i}, x_{2i}, s_i) \quad (1)$$

A $(2^{nR_{11}}, 2^{nR_{21}}, 2^{nR_{22}}, n, \epsilon)$ code for the discrete memoryless Z channel with side information consists of two sets of encoding mappings

$$e_1: \{1, 2, \ldots, 2^{nR_{11}}\} \times S^n \rightarrow X_1^n$$
$$e_2: \{1, 2, \ldots, 2^{nR_{21}}\} \times \{1, 2, \ldots, 2^{nR_{22}}\} \times S^n \rightarrow X_2^n$$

and two sets of decoding mappings

$$d_1: Y_1^n \rightarrow \{1, 2, \ldots, 2^{nR_{11}}\} \times \{1, 2, \ldots, 2^{nR_{21}}\}$$
$$d_2: Y_2^n \rightarrow \{1, 2, \ldots, 2^{nR_{22}}\}$$

and an average probability of error defined as the probability that the decoded message does not equal the transmitted message such that

$$p(d_1(y_1^n) \neq (m_{11}, m_{21}) \text{ or } d_2(y_2^n) \neq m_{22}) \leq \epsilon$$

where the messages are assumed to be uniformly distributed on their respective sets.

A rate triple $(R_{11}, R_{21}, R_{22})$ is said to be achievable for the discrete memoryless “Z” channel with side information if there exists a sequence of $(2^{nR_{11}}, 2^{nR_{21}}, 2^{nR_{22}}, n, \epsilon)$ codes.

III. THE MAIN RESULT

In this section, we derive an achievable rate region for the general Z channel with side information. At the first transmitter we apply the Gelfand-Pinsker random binning and at the second transmitter we use a combination of superposition coding, Marton encoding [7], Gelfand-Pinsker coding, and CMGE [8] jointly decoding.

**Definition 1:** Define $P_{ZCSI}$ as the set of all random variables $(S, W, X_1, U, U_2, X_2, Y_1, Y_2)$ such that

$$p(s,w,x_1,u,u_1,u_2,x_2,y_1,y_2) = p(s)p(w|s)p(x_1|w,s)p(u|s)p(u_1,u_2|u,s). \quad (2)$$

$$p(x_2|u,u_1,u_2,s)p(y_1,y_2|x_1,x_2,s)$$

where $(p(x_1|w,s), p(x_2|u,u_1,u_2,s)) \in [0,1]^2$

**Theorem 1:** An achievable rate region for the discrete memoryless Z channel with side information non-causally available at the transmitters, depicted in Fig. 2, is the closure of the convex hull of the set $R_{ZCSI} = U_{p \in P_{ZCSI}} R_{ZCSI}(p)$ where

$$R_{ZCSI}(p) = \{(R_{11}, R_{21}, R_{22}): R_{11} \leq A \quad (3-1)$$
$$R_{21} \leq B \quad (3-2)$$
$$R_{22} \leq C \quad (3-3)$$
$$R_{11} + R_{21} \leq D \quad (3-3)$$
$$R_{21} + R_{22} \leq \min(E,F) \quad (3-4)$$
$$R_{11} + R_{21} + R_{22} \leq \min(G,H,I) \quad (3-5)$$

for some $(S, W, X_1, U, U_2, X_2, Y_1, Y_2) \in P_{ZCSI}$

where we have
A = I(W; Y_1 | U_0, U_1) - I(W; S | U_0, U_1)
B = I(U_0, U_1; Y_1 | W) - I(U_0, U_1; S | W)
C = I(U_0, U_1; Y_2) - I(U_0, U_1; S)
D = I(W, U_0, U_1; Y_1) - I(W, U_0, U_1; S)
E = I(U_0, U_1; Y_1 | W) + I(U_2; Y_1 | U_0) - I(U_0, U_1; S | W)
F = I(U_1; Y_1 | W, U_0) + I(U_0, U_1; Y_2) - I(U_1, U_0; S | W)
G = I(W, U_1; Y_1 | U_0) + I(U_0, U_1; Y_2) - I(W, U_1; S | U_0)
H = I(W, U_0, U_1; Y_1) + I(U_2; Y_2 | U_0) - I(W, U_0, U_1; S)
I = I(W, U_0, U_1; Y_1) + I(U_0, U_1; Y_2) - I(W, U_0, U_1; S)

Corollary 1.1: If we put S ⊆ Φ in (3), then we have the achievable rate of the Z Channel provided by [4].

Corollary 1.2: If we let no information to be sent to the second receiver, we obtain the achievable rate for the state-dependent multiple access channel with independent sources, i.e. if we set R_20 = R_2 = 0, R_1 = R_2, U = U_2 = Φ, U_1 = U, W = U, and Y = Y in (21 - 33), we obtain the closure of the convex hull of all the state pairs (R_1, R_2) satisfying

R_1 ≤ I(U_1; Y | U_2) - I(U_1; S | U_2) (4)
R_2 ≤ I(U_1; Y | U_1) - I(U_1; S | U_1) (5)
R_1 + R_2 ≤ I(U_1; Y | U_1) - I(U_1; U_1; S) (6)

Corollary 1.3: If we set R_1 = R_2 = 0, R_20 = R_2, R_21 = R_1, and W = U_2 = Φ, U = U_2 in the 12 expressions derived from (21) - (33) and assuming that the receiver Y_2 is a degraded version of Y_1, we obtain the achievable rate region of the degraded broadcast channel with side information provided in [9].

R_1 ≤ I(U_1; Y | U) - I(U_1; S) (7)
R_2 ≤ I(U; Y_2) - (U; S) (8)

Proof: Fix a distribution of the form

p(s)p(w)p(x_1 | w)p(u_1)p(u_2 | u_1, w).

The second transmitter splits its messages as mentioned in section I. We then generate the codebook as follows

Randomly and independently generate 2^{n(R_11 + R_11)} sequences w^{n}(m_{11}, m_{11}) each one i.i.d according to \Pi_{i=1}^{n} p(w_i) and randomly partition them into 2^{nR_{11}} bins.

Randomly and independently generate 2^{n(R_21 + R_21 + R_{22})} sequences u^{n}(m_{21}, m_{22}, m_{21}, m_{22}) each one i.i.d according to \Pi_{i=1}^{n} p(u_i) and randomly partition them into 2^{nR_{22}} bins.

For each pair (m_{21}, m_{22}), independently generate 2^{n(R_{22} + R_{2k})} sequences u^{n}_{k}(m_{21}, m_{22}, m_{21}, m_{22}, m_{21}, m_{22}, m_{21}, m_{22}) with k = 1.2. each one i.i.d according to \Pi_{i=1}^{n} p(u_i) and randomly partition them into 2^{nR_{2k}} bins.

Encoding: Assume that the transmitters want to send the triple (m_{11}, m_{21}, m_{22}) with m_{2k} = (m_{2k}, m_{2k}).

TX1 looks in bin m_{11} to find some \tilde{m}_{11} such that (w^{n}(m_{11}, \tilde{m}_{11}), s^n) is jointly typical. Assume that the chosen index is \tilde{M}_{11}.

TX2, in the meantime, looks in bin (m_{21}, m_{22}) to find some pair (\tilde{m}_{21}, \tilde{m}_{22}) such that the pair (u^{n}(m_{21}, \tilde{m}_{21}, \tilde{m}_{22}, m_{21}, m_{22}, m_{2k}), s^n) is jointly typical. Assume that the chosen pair is (\tilde{M}_{21}, \tilde{M}_{22}).

TX2 then looks in bin m_{2k} to find some \tilde{m}_{2k} such that the pair (u^{n}_{k}(m_{21}, \tilde{m}_{21}, \tilde{m}_{22}, m_{21}, m_{22}, m_{2k}, m_{2k}), s^n) is conditionally jointly typical given u^n. Given that TX2 has found some \tilde{m}_{2k}, that satisfy the above condition, it looks in bins (m_{21}, m_{22}) to find some (u^n_{1}, u^n_{2}) such that the tuple

(u^n, u^n_{1}, u^n_{2}, s^n)

is jointly typical.

TX1 and TX2 then send

x_{11} = x_{11}(w_{11}, s_{1})
x_{21} = x_{21}(u_{11}, u_{12}, u_{21}, s_{1})

Decoding: Without loss of generality, assume that the triple (1,1,1) was sent through the channel. The first receiver receives y^n_{1} and looks for the unique pair (\tilde{m}_{11}, \tilde{m}_{21}) such that

(w^{n}(\tilde{m}_{11}, \tilde{m}_{11}), u^{n}(\tilde{m}_{21}, \tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{22}, \tilde{m}_{21}), y^n_{1}) \in A^{(n)}

where A^{(n)} is the set of jointly typical sequences.

The second receiver, meanwhile, receives y^n_{2} looks for the unique message index \tilde{m}_{22} such that

(u^n(\tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{22}, \tilde{m}_{22}, \tilde{m}_{21}, \tilde{m}_{21}), y^n_{2}) \in A^{(n)}

We define the following error events for the encoding section

E_{1}^{enc} = \{(w^{n}(m_{11}, \tilde{m}_{11}), s^n) \notin A^{(n)} for all \tilde{m}_{11} \in [1, 2^{nR_{11}}]\}
E_{2}^{enc} = \{(u^{n}(m_{21}, m_{22}, \tilde{m}_{21}, \tilde{m}_{22}), s^n) \notin A^{(n)} for all \tilde{m}_{21}, \tilde{m}_{22} \in [1, 2^{nR_{21} \times 1, 2^{nR_{22}}}]\}
E_{3k}^{enc} = \{(u^{n}(m_{21}, m_{22}, \tilde{m}_{21}, \tilde{m}_{22}), u^{n}_{k}(m_{21}, m_{22}, \tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{21}, \tilde{m}_{22}, m_{2k}), s^n) \notin A^{(n)} for all \tilde{m}_{2k} \in [1, 2^{nR_{2k}}] \} \}, k = 1, 2.
E_{4}^{enc} = \{(u^{n}(m_{21}, m_{22}, \tilde{m}_{21}, \tilde{m}_{22}), u^{n}_{k}(m_{21}, m_{22}, \tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{21}, \tilde{m}_{22}), s^n) \notin A^{(n)} for all \tilde{m}_{2k} \in [1, 2^{nR_{2k}}] \} \cup [1, 2^{nR_{2k}}] \}

The decoding error events for the first receiver are defined as follows
In just the same way one can prove that
\[
e^{-n((W,S) + \delta_1(v) - \tilde{R}_{11})} \leq \frac{\epsilon}{13}
\]
provided that
\[
\tilde{R}_{11} \geq I(W;S) + \delta_1(\epsilon)
\]
In the same way we can prove that
\[
p(E_1^{enc}) + p(E_2^{enc}) + p(E_3^{enc}) \leq \frac{4\epsilon}{13}
\]
provided that
\[
\tilde{R}_{21}^e + \tilde{R}_{22}^e \geq I(U;S) + \delta_2(\epsilon)
\]
\[
\tilde{R}_{21}^e \geq I(U_1;S|U) + \delta_3(\epsilon)
\]
\[
\tilde{R}_{22}^e \geq I(U_2;S|U) + \delta_4(\epsilon)
\]
The probability of error at the first receiver is bound as follows
\[
p(E_1^{dec}) = \frac{\sum_{(m_{11},m_{12},m_{21},m_{22})=1}^{2^{nR_{11}}} \sum_{(u^n, u^n_r, u^n_{12})=1}^{2^{nR_{11}}} p(u^n) p(u^n_r) p(u^n_{12})}{2^n (R_{11} + R_{21}^e + R_{22}^e + R_{31}^e + R_{32}^e + R_{41}^e + R_{42}^e + R_{51}^e) = I(WU;U_1|Y_1) + I(W;U) + I(W;U_1|U) - \Sigma R_{11} - 7\epsilon} \leq \frac{\epsilon}{13}
\]
provided that
\[
R_{11} + \tilde{R}_{11} + R_{21}^e + R_{22}^e + R_{31}^e + R_{32}^e + R_{41}^e + R_{42}^e \leq I(WU;U_1|Y_1) + I(W;U) + I(W;U_1|U) - 7\epsilon
\]
where
\[
\sum_{R_{11} = 1}^{6} R_{11} + \tilde{R}_{11} + R_{21}^e + R_{22}^e + R_{31}^e + R_{32}^e + R_{41}^e + R_{42}^e + R_{51}^e \leq I(WU;U_1|Y_1) + I(W;U) + I(W;U_1|U) - 7\epsilon
\]
In just the same way one can prove that
\[
\sum_{k=3}^{6} p(E_k^{dec}) \leq \frac{4\epsilon}{13}
\]
provided that
\[
R_{21}^e + R_{22}^e + R_{31}^e + R_{32}^e + R_{41}^e + R_{42}^e \leq I(UU_1;Y_1|W) + I(W;U) + I(W;U_1|U)
\]
\[
R_{11} + \tilde{R}_{11} + R_{21}^e + R_{22}^e \leq I(WU_1;Y_1|U) + I(W;U) + I(W;U_1|U)
\]
\[
R_{21}^e + R_{22}^e \leq I(U_1;Y_1|UW) + I(W;U) + I(W;U_1|U)
\]
\[ R_{11} + R_{12} \leq I(W; Y_1 | U_1) + I(W; U_1 | Y_1) + I(W; U_1 | U_1) \]  
\[ \text{(18)} \]

For the second receiver, the analysis of error events imply that
\[ p(E_{21}^{\text{dec}}) + p(E_{22}^{\text{dec}}) \leq \frac{2e}{13} \]

provided that
\[ R_{21}^p + R_{22}^p + R_{23}^p + R_{24}^p + R_{25}^p \leq I(UU_2; Y_2) \]  
\[ R_{22}^p + R_{23}^p \leq I(U_2; Y_2 | U) \]  
\[ \text{(19)} \]
\[ \text{(20)} \]

Therefore, the probability of error can be bounded as
\[ P_e^{(n)} = p \left( \left( \bigcup_{k=1}^{4} E_{k1}^{\text{enc}} \right) \cup \left( \bigcup_{k=1}^{6} E_{1k}^{\text{dec}} \right) \cup \left( \bigcup_{k=1}^{2} E_{2k}^{\text{dec}} \right) \right) \leq 5e + 6e + 2e \]
\[ \frac{13}{13} + \frac{13}{13} = \epsilon \]

Now combining (14) – (20) with (9) – (13), and setting
\[ R_{50} = R_{21} + R_{22} \]
we obtain the following expressions
\[ R_{11} \leq A \]  
\[ R_{11} + R_{21}^p \leq B \]  
\[ R_{11} + R_{22}^p + R_{23}^p \leq C \]  
\[ R_{11} + R_{23}^p + R_{24}^p \leq D \]  
\[ R_{11} + R_{23}^p + R_{24}^p \leq \min\{E, F\} \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq G \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq H \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq I \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq J \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq K \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq L \]  
\[ R_{11} + 2R_{20} + R_{21} + R_{22}^p \leq \min\{M, N\} \]  
\[ 2R_{20} + R_{21} + R_{22}^p \leq O \]  
\[ \text{(21)} \]
\[ \text{(22)} \]
\[ \text{(23)} \]
\[ \text{(24)} \]
\[ \text{(25)} \]
\[ \text{(26)} \]
\[ \text{(27)} \]
\[ \text{(28)} \]
\[ \text{(29)} \]
\[ \text{(30)} \]
\[ \text{(31)} \]
\[ \text{(32)} \]
\[ \text{(33)} \]

where
\[ A = I(Y_1; W | U_1) + I(W; U, U_1) - I(W; S) \]
\[ B = I(Y_1; W, U_1 | U) + I(W; U, U_1) - I(U_1; S | U) - I(W; S) \]
\[ C = I(Y_1; W, U, U_1) + I(W; U, U_1) - I(U, U_1; S) - I(W; S) \]
\[ D = I(Y_1; W, U_1 | U) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(W; S) - I(U_1, U_2 | S | U) \]
\[ E = I(Y_1; W, U_1 | U) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(W; S) - I(U_1, U_2 | S | U) \]
\[ F = I(Y_1; W, U_1 | U) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(W; S) - I(U_1, U_2 | S | U) \]
\[ G = I(Y_1; W | UU_1) + I(Y_2; U_2 | U) + I(W; UU_1) \]
\[ - I(U_1; U_2 | U) - I(W; S) - I(U_1, U_2 | S | U) \]
\[ H = I(Y_1; U_1 | W, U) + I(W; U, U_1) - I(U_1; S | U) \]
\[ I = I(Y_1; U_1 | U) - I(U_2; S | U) \]
\[ J = I(Y_1; U_1 | W, U) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(U_1, U_2 | S | U) \]
\[ K = I(Y_1; U_1 | W) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(U_1, U_2 | S | U) \]
\[ L = I(Y_1; U_1 | U) - I(U_2; S | U) \]
\[ M = I(Y_1; U_1 | W) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(U_1, U_2; S) \]
\[ N = I(Y_1; U_1 | W, U) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(U_1, U_2; S) \]

\[ O = I(Y_1; U_1 | W) + I(Y_2; U_2 | U) + I(W; U, U_1) \]
\[ - I(U_1; U_2 | U) - I(U; S) - I(U, U_1, U_2; S) \]

The expressions (21) – (33) first undergo a Fourier-Motzkin procedure using lemma 1 in [16] after which we will have 12 expressions exactly the same as (21) – (33) with (27) – (29) removed. Now applying the Fourier-Motzkin elimination scheme to the 12 expressions derived from the last step using the constraints \( R_{21} + R_{22} = R_{20} + R_{21} + R_{22}^p, R_{21}^2 \leq R_{21}, R_{22}^p \leq R_{22}, \) and nonnegativity of the rates, we obtain the inequalities in (3).
\[ R_{22} \leq I(U_2;Y_2|U) - I(U_2;S|U) \]  

(42)

for some distribution of the form \( p(s,w,x_1,u,u_2,x_2) = p(s)p(w|s)p(x_1|w,s)p(u|s)p(u_2|u_2,s)p(x_2|u, u_2, s) \) constitutes an outer bound on the capacity region of the degraded discrete memoryless \( Z \) channel with side information.

**Proof:** See Appendix II.

**Remark 3.1:** Notice that achievable rates of the second transmitter coincide with their counterparts in the outer bound optimally with the receivers and therefore, the second transmitter can communicate optimally with the receivers.

**C. Achievable rate for the degraded Gaussian \( Z \) channel with interference**

Now we study the Gaussian version of the \( Z \) channel with channel state information. First we define the Gaussian \( Z \) channel model with interference. Then we evaluate the achievable rate found for the discrete memoryless degraded \( Z \) channel with side information to the Gaussian case and use dirty paper coding to remove the negative effect of the interference in the channel associated with the first transmitter-receiver pair. The general model of the Gaussian \( Z \) channel is as follows:

\[
\begin{align*}
Y_1 &= X_1 + aX_2 + a_1S + Z_1 \\
Y_2 &= X_2 + a_2S + Z_2
\end{align*}
\]  

(43)  

(44)

where for \( k = 1,2 \)

\[ \frac{1}{n} \sum_{i=1}^{n} E(X_k)^2 \leq P_k, Z_k \sim N(0, N_k), \text{ and } S \sim N(0, Q). \]

Now we use the dirty paper coding presented in [11] to derive the Gaussian version of the achievable rate presented in Theorem 2. We first present a Lemma to prove that the Gaussian version of Theorem 2, i.e. Theorem 4, is indeed the achievable rate found for the discrete memoryless degraded \( Z \) channel with side information to the Gaussian case and use dirty paper coding to remove the negative effect of the interference in the channel associated with the first transmitter-receiver pair. The general model of the Gaussian \( Z \) channel is as follows:

\[
\begin{align*}
Y_1 &= X_1 + aX_2 + a_1S + Z_1 \\
Y_2 &= X_2 + a_2S + Z_2
\end{align*}
\]  

(43)  

(44)

\[
\begin{align*}
I(U;Y_i|W) &= I(U;Y_i,S|W) \\
I(W;Y_i|U) &= I(W;Y_i,S|U)
\end{align*}
\]  

(46)

provided that

\[
\begin{align*}
\alpha &= \frac{aa_1\sqrt{N_2}}{N_1 + P_1 + a^2P_2} \\
\beta &= \frac{a_1\sqrt{P_1}}{N_1 + P_1 + a^2P_2}
\end{align*}
\]  

(47)  

(48)

**Proof:** Using the definitions in (45), the equalities in (46), and the inequalities (34) – (38), we derive (66) – (70).

**Remark 4.1:** Notice that with the Costa-coefficients of (47) and (48), we can remove the effect of the interference from the first three inequalities. In fact, it can be shown that (47) changes if one desires to remove the effect of interference in the bounds on the rates of the second transmitter.

**Remark 4.2:** Notice that inequalities (66) – (70), are like those found in Corollary 1 of [3] where there is no interference.

**V. Lattice Strategies for the Gaussian Degraded “\( Z \)” Channel with Additive Interference**

Now we propose an achievable rate region using lattice based coding for the Gaussian degraded “\( Z \)” channel with additive interference non-causally available at both of the encoders under the high-SNR and strong interference regime utilizing the standard notation of [13], [14], and [15]. Our method is based on lattice transmission scheme, jointly decoding at the first decoder and successive decoding at the second decoder. Exploiting such coding scheme we remove the effect of the interference completely. The model of the Gaussian degraded Z channel with the additive interference that we use in this section is depicted in Fig. 4.

The outputs are:

\[
\begin{align*}
Y_1 &= X_1 + aX_2 + (1 + \alpha)S + Z_1 \\
Y_2 &= X_2 + S + Z_2
\end{align*}
\]  

(49)  

(50)

where \( \alpha \) is a real number, \( X_i, i = 1,2 \), is the channel input transmitted by user \( i \) which is subject to the power constraint \( P_i \), \( Z_i \) is an AWGN with zero mean and variance \( N_i \) \((Z_i \sim N(0, N_i))\), and the interference signal \( S \) is assumed to be i.i.d. Gaussian with variance \( Q \), i.e., \( S \sim N(0, Q) \), independent of everything else and known non-causally at both encoders.
Theorem 5: An achievable region rate for the Gaussian degraded “Z” channel with side information non-causally available at the transmitters, denoted by $\mathcal{R}$, is given by

$$\mathcal{R} = \bigcup_{\rho \in [0,1] \cup \mathbb{R}^+} \mathcal{R}(\rho, \alpha_0)$$

where

$$A_1 = \frac{1}{2} \log \left( \frac{\alpha_0^2 \rho P_2}{\alpha_0^2 \rho P_2 + (\alpha_0^2 \rho P_2 + N_1)} \right) \times \left( \frac{\alpha_0^2 \rho P_2 + N_1 + P_1}{\alpha_0^2 \rho P_2 + N_1} \right) \times \left( \frac{\alpha_0^2 \rho P_2 + N_1 + P_1}{\alpha_0^2 \rho P_2 + N_1} \right)$$

$$A_2 = \frac{1}{2} \log \left( \frac{\rho P_2}{\alpha_0^2 \rho P_2 + N_2} \right) \times \left( \frac{\rho P_2 + N_1 + P_1}{\rho P_2 + N_1 + P_1} \right) \times \left( \frac{\rho P_2 + N_1 + P_1}{\rho P_2 + N_1 + P_1} \right)$$

Proof: Our method is based on lattice transmission scheme, jointly decoding at the first decoder and successive decoding at the second decoder.

Encoding: Consider three lattices $\Lambda_i$, $i = 0,1,2$ with fundamental Voronoi regions $V_i$, and the second moments $\sigma_{\Lambda_i}^2 = \rho P_2, \sigma_{\Lambda_0}^2 = P_1$ and $\sigma_{\Lambda_2}^2 = \bar{\rho} P_2$, respectively. Encoder 1 wants to send message $m_1$ to receiver 1 while encoder 2 wants to send message $m_0$ to both receivers and message $m_2$ to receiver 2. We let $R_0 = R_{21}, R_1 = R_{11}$ and $R_2 = R_{22}$. The messages $m_i$ are carried by vector $V_i$ where $V_i$ is uniformly distributed over $V_i$, and $V_0, V_1$ and $V_2$, are pair-wise independent. Also, let $D_i \sim \text{Unif}(V_i), i = 0,1,2$, be three dither signals which are uniformly distributed over $V_i$, and independent of each other. In our encoding structure senders 1 and 2 send $X_1 = W$ and $X_2 = U + U_2$, respectively, where $U, W$ and $U_2$ are generated as:

$$U = [V_0 - \alpha_0 S + D_0] \mod \Lambda_0$$

$$W = [V_1 - \alpha_0 \bar{\alpha}_0 S + D_1] \mod \Lambda_1$$

$$U_2 = [V_2 - \alpha_2 \bar{\alpha}_0 S + D_2] \mod \Lambda_2$$

where $\bar{\alpha}_0 = 1 - \alpha_0$ and the MMSE criterion is used to determine $\alpha_i, i = 0,1,2$. Note that using this encoding structure we have: $\frac{1}{n} E[\|X_1\|^2] = P_1$ and $\frac{1}{n} E[\|X_2\|^2] = P_2$.

Decoding: To decode $(V_0, V_2)$ at decoder 2, we use a successive decoding scheme in which decoder 2 first decodes $V_0$ and then decodes $V_2$. Therefore, decoder 2, after receiving $Y_2$ and using the lattice $\Lambda_0$, computes

$$Y_0^{(2)} = [\alpha_0 Y_2 - D_0] \mod \Lambda_0$$

$$= [\alpha_0 (U + U_2 + S + Z_2) - D_0] \mod \Lambda_0$$

$$= [V_0 - \bar{\alpha}_0 U + \alpha_0 (U_2 + Z_2)] \mod \Lambda_0$$

$$= [V_0 + Z_{02,x}] \mod \Lambda_0$$

where $Z_{02,x} = -\bar{\alpha}_0 U + \alpha_0 (U_2 + Z_2)$. Therefore, we have:

$$R_0 = \frac{1}{n} I(V_0; Y_0^{(2)}) = \frac{1}{n} \left\{ h(Y_0^{(2)}) - h(Y_0^{(2)}|V_0) \right\}$$

$$\leq \frac{1}{2} \left( \rho P_2 \right) - \frac{1}{2} \left( 2 \pi e \left( \alpha_0^2 \rho P_2 + \alpha_0 \rho \bar{\rho} P_2 + N_2 \right) \right)$$

Note that using this encoding $\alpha_0^2 \rho P_2 + \alpha_0 \rho P_2 + N_2$ is non-causal. Therefore, we have:

$$R_0 \leq \frac{1}{2} \left( 1 + \frac{\rho P_2}{\rho P_2 + \rho \bar{\rho} P_2 + N_2} \right)$$

Also, note that this $\alpha_0^2 \rho P_2 + \alpha_0 \rho \bar{\rho} P_2 + N_2$ is non-optimal from the first receiver standpoint.

Now, decoder 2 using lattice $\Lambda_2$ computes

$$Y_2^{(2)} = [\alpha_2 (\bar{\alpha}_0 Y_2 + Z_{02,x}) - D_2] \mod \Lambda_2$$

$$= [\alpha_2 (U_2 + Z_2) + \alpha_2 \bar{\alpha}_0 S - D_2] \mod \Lambda_2$$

$$= [V_2 - \bar{\alpha}_2 U_2 + \alpha_2 Z_2] \mod \Lambda_2$$

Therefore, we have:

$$R_2 = \frac{1}{n} I(V_2; Y_2^{(2)}) = \frac{1}{n} \left\{ h(Y_2^{(2)}) - h(Y_2^{(2)}|V_2) \right\}$$

$$\leq \frac{1}{2} \left( 1 + \frac{\rho P_2}{\rho P_2 + \rho \bar{\rho} P_2 + N_2} \right)$$

Also, note that this $\alpha_2^2 \rho P_2 + \alpha_2 \rho \bar{\rho} P_2 + N_2$ is non-optimal from the first receiver standpoint.
\[
\geq \frac{1}{2} \log \left( \frac{\tilde{P}_2}{\mathcal{G}(\Lambda_2)} \right) - \frac{1}{2} \log \left( 2\pi e \left( \tilde{a}_2 \tilde{\rho}_2 + \alpha_2^2 N_2 \right) \right) 
\]

and for good lattice for quantization, the achievable rate is given by

\[
R_2 \leq \frac{1}{2} \log \left( \frac{\tilde{P}_2}{a_2^2 \tilde{\rho}_2 + \alpha_2^2 N_2} \right) 
\]

Note that the optimal \( \alpha_2 \) is \( \alpha_2^{\text{opt}} = \frac{\tilde{\rho}_2}{\tilde{\rho}_2 + a_2^2 N_2} \), and by substituting this \( \alpha_2^{\text{opt}} \) into (60), we obtain:

\[
R_2 \leq \frac{1}{2} \log \left( 1 + \frac{\tilde{P}_2}{N_2} \right) 
\]  (61)

To decode \((V_0, V_1)\) at decoder 1, we use a similar method as [9] for MAC. Therefore, as long as \( \Lambda_0 \) is a good lattice for quantization, we have:

\[
R_0 \leq \frac{1}{2} \log \left( \frac{a_0^2 \rho_2}{a_0^2 a_2^2 \rho_2 + a_0^2 (a_2^2 \rho_2 + N_1)} \right) = A_1 
\]  (62)

Note that the optimal \( \alpha_0 \) for sender 2 from the first receiver standpoint is \( \alpha_0^{\text{opt}} = \frac{\rho_2}{a_2^2 \rho_2 + a_2^2 N_1} \), and by substituting this \( \alpha_0^{\text{opt}} \) into (62), we obtain:

\[
R_0 \leq \frac{1}{2} \log \left( 1 + \frac{a_0^2 \rho_2}{a_0^2 a_2^2 \rho_2 + N_1} \right) 
\]  (63)

Also, note that this \( \alpha_0^{\text{opt}} \) is non-optimal from the second receiver standpoint. Similarly, for good lattice for quantization we have:

\[
R_1 \leq \frac{1}{2} \log \left( \frac{P_1}{\tilde{a}_1 \tilde{\rho}_1 + \tilde{a}_1^2 (a_2^2 \rho_2 + N_1)} \right) 
\]  (64)

Meanwhile, the optimal \( \alpha_1 \) is \( \alpha_1^{\text{opt}} = \frac{P_1}{P_1 + a_2^2 \tilde{\rho}_2 + N_1} \), and by substituting this \( \alpha_1^{\text{opt}} \) into (64), we obtain:

\[
R_1 \leq \frac{1}{2} \log \left( 1 + \frac{P_1}{a_2^2 \tilde{\rho}_2 + N_1} \right) 
\]  (65)

VI. CONCLUSION

In this paper, we derived an achievable rate region for the general “Z” channel with side information non-causally available at the transmitters using Han-Kobayashi, rate splitting, Gelfand-Pinsker coding, and CMGE jointly decoding. We also showed that our rate region subsumes the achievable rate region of the multiple access channel with side information and degraded broadcast channel with side information. We then derived the achievable rate region of a special case of degraded Z channels with side information. We also saw that using the Costa dirty paper coding for the degraded Gaussian Z channel, we can remove the effect of interference in the direction of one of the transmitter-receiver pairs. We saw that using dirty paper coding, we can remove the effect of the interference completely.

REFERENCES

[1] S. Vishwanath, N. Jindal, and A. Goldsmith, “The “Z” channel,” IEEE Global Telecommunications Conference, pp. 1726-1730, Dec. 2003.
[2] N. Liu, S Ulukus, “On the capacity region of the Gaussian Z channel,” IEEE Global Telecommunications Conference, pp. 415-419, 29 Nov.-3 Dec. 2004.
[3] Hon-Fah Chong, Mehul Motani, and Hari Krishna Garg, “Capacity Theorems for the “Z” channel,” IEEE Trans. on Inf. Theory, vol. 53, pp. 1348-1365, April 2007.
[4] H. T. Do, T. J. Oechtering, and M. Skoglund, “Capacity bounds for the Z channel,” IEEE Info. Theory Workshop, pp. 432-436, Oct. 2011.
[5] C. E. Shannon, “Channels with Side Information at the Transmitter,” IBM Journal of Research and Development, Vol. 2, pp. 289-293, Oct. 1958.
[6] S. I. Gel’fand and M. S. Pinsker, “Coding for Channel with random parameters,” Probl. Contr. And Inform. Theory, Vol. 9, no. 1, pp. 19-31, 1980.
[7] T. Han, and K. Kobayashi, “A new achievable rate region for the interference channel,” IEEE Trans. on Inf. Theory, pp. 49-60, vol. 27, Jan. 1981.
[8] H. F. Chong, M. Motani, H. K. Garg, H. El Gamal, “On the Han-Kobayashi region for the interference channel,” IEEE Trans. on Inf. Theory, pp. 3188-3195, vol. 54, July 2008.
[9] Y. steinberg, “Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information,” IEEE Trans. on Inf. Theory, pp. 2867-2877, vol. 51, Aug. 2005.
[10] A. Carleial, “Interference channels,” IEEE Trans. on Inf. Theory, pp. 60-70, vol. 24, Jan. 1978.
[11] M. Costa, “Writing on dirty paper,” IEEE Trans. on Inf. Theory, pp. 439-441, vol. 29, May 1983.
[12] I. Csiszár, J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Budapest: Akademiai Kiado, 1997.
[13] T. Philosof, R. Zamir, U. Erez, and A.J. Khisti, “Lattice strategies for the dirty multiple access channel,” IEEE Trans. Inf. Theory, vol. 57, pp. 5006–5035, Aug. 2011.
[14] R. Zamir and M. Feder, “On lattice quantization noise,” IEEE Trans. Inf. Theory, pages 1152–1159, July 1996.
[15] G. D. Forney Jr., “On the role of MMSE estimation in approaching the information theoretic limits of linear Gaussian channels: Shannon meets Wiener,” in Proc. Allerton Conf. Commun., Control and Comp., 2003.
[16] Y. Liang, G. Kramer, “Rate Regions for Relay Broadcast Channels,” IEEE Trans. Info. Theory, pages 3517–3535, Oct. 2007.
[17] C. Nair, Z. V. Wang, “The capacity region of the three receiver less noisy broadcast channel,” IEEE Trans. Inf. Theory, pp. 4058-4062, vol. 57, no. 7 , July 2011.
In fact, it suffices to show that

First we prove that

\[ I(UW; Y_1) = I(UW; Y_1, S) \]

In fact, it suffices to show that

\[ h(UW | Y_1) = h(UW | Y_1, S). \]

We have

\[ h(UW | Y_1) = h(U \mid aS, \overline{W} + \beta S | X_1 + aX_2 + a_1S + Z_1) \]

\[ = h \left( \overline{U} - \frac{\alpha}{a_1} \left( \sqrt{P_1 \overline{W} + a_1 \sqrt{P_2 \overline{U}}} + a \sqrt{P_2 \overline{U}_2 + Z_1} \right) \right) \]

\[ = h \left( \psi_w, \psi_w \right) \]

\[ = \begin{align*}
\psi_w \times \left( \sqrt{P_1 \overline{W} + a_1 \sqrt{P_2 \overline{U}}} + a \sqrt{P_2 \overline{U}_2 + Z_1} \right) = E \left[ \psi_w \times \left( \sqrt{P_1 \overline{W} + a_1 \sqrt{P_2 \overline{U}}} + a \sqrt{P_2 \overline{U}_2 + Z_1} \right) \right] = 0
\end{align*} \]

provided that

\[ \alpha = \frac{a a_1 \sqrt{P_2}}{N_1 + P_1 + a^2 P_2} \]
The other equalities of (46) are proved in the same way.

APPENDIX II

Before proving Theorem 3, we need to prove a lemma which will be needed throughout the course of the proof procedure.

**Lemma 2:** Let the discrete memoryless channel \( X \rightarrow Y \) be a less noisy version of the discrete memoryless channel \( X \rightarrow Y \).

Consider \((M, S^{i-1})\) for every \( i = 1, 2, \ldots, n \) to be any random vector underlying the state-dependent broadcast channel \( X \rightarrow (Y_i, Y_{2i}) \) and also from the discrete memorylessness of the channel

\[
(M, S^{i-1}) \rightarrow (X_i, S_i) \rightarrow (Y_{i, i}, Y_{2i, i})
\]

Forms a Markov chain. Then

\[
I(Y_{i+1}^{(n)}; Y_i | M, S^{i-1}) \leq I(Y_{2i+1}^{(n)}; Y_i | M, S^{i-1})
\]

**Proof:** For any \( i + 1 \leq r \leq n - 1 \) we have

\[
I(Y_{i+1}^{(n)}, Y_{2r+1}^{(n)}; Y_i | M, S^{i-1}) = I(Y_{i+1}^{(n)}, Y_{2r+1}^{(n)}; Y_i | M, S^{i-1}) + I(Y_i; Y_i | M, S^{i-1}, Y_{i+1}^{(n)}, Y_{2r+1}^{(n)})
\]

\[
\leq I(Y_{i+1}^{(n)}, Y_{2r+1}^{(n)}; Y_i | M, S^{i-1}) + I(Y_{2r+1}; Y_i | M, S^{i-1}, Y_{i+1}^{(n)}, Y_{2r+1}^{(n)})
\]

\[
= I(Y_{i+1}^{(n)}, Y_{2r+1}^{(n)}; Y_i | M, S^{i-1})
\]

where \((a)\) follows from the fact that

\[
(M, S^{i-1}, Y_{i+1}^{(n)}, Y_{2r+1}^{(n)}, Y_i) \rightarrow (X_r, S_r) \rightarrow (Y_{i, i}, Y_{2i, i})
\]

and also from the fact that \( Y_2 \) is a less noisy version of \( Y_1 \). Applying the above inequality a number of times yields the Lemma.

**Proof of Theorem 3:** Suppose that \( (2^{nR_1}, 2^{nR_{21}}, 2^{nR_{22}}, n, \epsilon) \) is a code for the degraded discrete memoryless Z channel. We define the auxiliary random variables as follows

\[
W_i \triangleq (M_1, S_{i+1}^{(n)})
\]

\[
U_j \triangleq (M_{2i}, Y_{2i+1}^{(n)}, S^{i-1})
\]

\[
U_{2i} \triangleq (M_{2i}, M_{22}, Y_{2i+1}^{(n)}, S^{i-1})
\]

First we prove the bound on \( R_{11} + R_{21} \). We have

\[
n(R_{11} + R_{21}) \leq H(M_1, M_{21} | Y_1^{(n)}) + I(M_1, M_{21}; Y_1^{(n)}) - I(M_1, M_{21}; S^{n})
\]

\[
\leq n \epsilon_{11n} + \sum_{i=1}^{n} I(M_{1i}, M_{21}; Y_{1i} | Y_{1i+1}^{(n)}) - I(M_{1i}, M_{21}; S^{i-1})
\]

\[
\leq n \epsilon_{11n} + \sum_{i=1}^{n} I(M_{1i}, M_{21}; Y_{1i} | Y_{1i+1}^{(n)}) - I(M_{1i}, M_{21}; S^{i-1})
\]

\[
= n \epsilon_{11n} + \sum_{i=1}^{n} I(M_{1i}, M_{21}; Y_{1i} | Y_{1i+1}^{(n)}) - I(S^{i-1}; Y_{1i} | M_{1i}, M_{21}, Y_{1i+1}^{(n)})
\]
where (a) follows from the independence of the messages from the state of the channel, (b) follows from Fano’s inequality and the chain rule for mutual information, (c) follows from the non-negativity of mutual information and from the fact that the channel state elements are i.i.d., (d) follows from Csiszar-Körner identity [12], (e) follows from Lemma 2 and non-negativity of mutual information, (f) follows from the i.i.d-ness of channel state, (g) follows from the two ways that $I(M_{11}; S^n)$ can be extended, and (h) follows from the i.i.d-ness of channel state and from non-negativity of mutual information.

Next, we prove the bound on $R_{21}$. We have

\[
nR_{21} = H(M_{21}|M_{11}, S^n, Y^n) + I(M_{21}; Y^n|M_{11}, S^n) \\
\leq n\varepsilon_{12n} + \sum_{i=1}^{n} I(M_{21}; Y^n_{i+1}|M_{11}, S^n_{i+1}, Y^n_{i+1}) \\
\leq n\varepsilon_{12n} + \sum_{i=1}^{n} I(M_{21}, S^n_{i+1}, Y^n_{i+1}|M_{11}, S^n_i) \\
\leq n\varepsilon_{12n} + \sum_{i=1}^{n} I(M_{21}, Y^n_{i+1}|M_{11}, S^n_i) + I(Y^n_{2i+1}|Y^n_{i+1}) \\
\leq n\varepsilon_{12n} + \sum_{i=1}^{n} I(U_i; Y^n_{i+1}|S^n_i, W_i) \\
= n\varepsilon_{12n} + \sum_{i=1}^{n} I(U_i; Y^n_{i+1}|S^n_i, W_i)
\]

where (a) follows from Lemma 2.

Now we prove the bound on $R_{21} + R_{22}$. We have

\[
n(R_{21} + R_{22}) = H(M_{21}, M_{22}|Y^n) + I(M_{21}, M_{22}; Y^n) - I(M_{21}, M_{22}; S^n) \\
\leq n\varepsilon_{21n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, Y^n_{2i+1}, Y^n_{2i}) - I(M_{21}, M_{22}, S^n_{i-1}; S_i)
\]
\[ n \epsilon_{21n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, Y_{2i+1}^n, S_{i-1}; Y_{2i}) - I(S_{i-1}; Y_{2i} | M_{21}, M_{22}, Y_{2i+1}^n) - I(M_{21}, M_{22}, S_{i-1}, Y_{2i+1}^n; S_i) + I(Y_{2i+1}^n; S_i | M_{21}, M_{22}, S_{i-1}) \]

\[
= (a) n \epsilon_{21n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, Y_{2i+1}^n, S_{i-1}; Y_{2i}) - I(M_{21}, M_{22}, S_{i-1}, Y_{2i+1}^n; S_i) + I(U_i, U_{2i}; Y_{2i}) - I(U_i, U_{2i}; S_i).
\]

where (a) follows from Csiszar-Körner identity [12].

Finally, we prove the bound on \( R_{22} \). We have

\[
n R_{22} = H(M_{22} | M_{21}, Y_2^n) + I(M_{22}; Y_2^n | M_{21}) - I(M_{22}; S^n | M_{21})
\]

\[
\leq n \epsilon_{22n} + \sum_{i=1}^{n} I(M_{22}; Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) - I(M_{22}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(M_{22}, S_{i-1}, Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) - I(M_{22}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, S_{i-1}, Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) - I(M_{22}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, S_{i-1}, Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) - I(M_{22}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(S_{i-1}, Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) + I(M_{21}, M_{22}, Y_{2i+1}^{(n)}; Y_{2i} | M_{21}, Y_{2i+1}^{(n)}, S_{i-1}) - I(M_{21}, M_{22}, S_{i-1}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(S_{i-1}, Y_{2i} | M_{21}, Y_{2i+1}^{(n)}) + I(M_{21}, M_{22}, Y_{2i+1}^{(n)}; Y_{2i} | M_{21}, Y_{2i+1}^{(n)}, S_{i-1}) - I(M_{21}, M_{22}, S_{i-1}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(M_{21}, M_{22}, Y_{2i+1}^{(n)}; Y_{2i} | M_{21}, Y_{2i+1}^{(n)}, S_{i-1}) - I(M_{21}, M_{22}, S_{i-1}, Y_{2i+1}^{(n)}; S_{i} | M_{21}, S_{i-1})
\]

\[
= n \epsilon_{22n} + \sum_{i=1}^{n} I(U_{2i}; Y_{2i} | U_i) - I(U_{2i}; S_i | U_i).
\]

where (a) and (b) both follow from Csiszar-Körner identity [12].