An Achievability Scheme for the Compound Channel with State Noncausally Available at the Encoder

Chandra Nair*, Abbas El Gamal† and Yeow-Khiang Chia†

Abstract

A new achievability scheme for the compound channel with discrete memoryless (DM) state noncausally available at the encoder is established. Achievability is proved using superposition coding, Marton coding, joint typicality encoding, and indirect decoding. The scheme is shown to achieve strictly higher rate than the straightforward extension of the Gelfand-Pinsker coding scheme for a single DMC with DM state, and is optimal for some classes of channels.

I. Introduction

Consider the problem of reliable communication over a compound channel with discrete memoryless (DM) state, where a sender wishes to communicate a message to a receiver with the state sequence available noncausually at the encoder. For simplicity we consider the case when the compound channel comprises only two discrete memoryless channels (DMCs) with DM state. This setup is essentially the same as sending a common message over a 2-receiver discrete memoryless broadcast channel (DM-BC) with DM state when the state in available noncausally at the encoder as shown in Figure 1. As such, we focus our discussion throughout the paper on this equivalent setup.

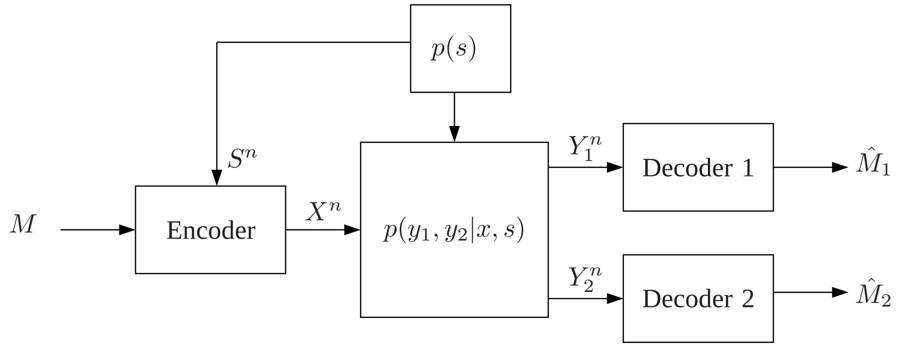

Fig. 1: Sending common message over DM-BC with DM state.

The capacity for the single receiver case, widely referred to as the Gelfand–Pinsker channel, was established in [1] as

$$C_{\mathrm{GP}} = \max_{p(u|s),\, x(u,s)} (I(U:Y) - I(U;S)).$$

The proof of achievability involves randomly generating a subcodebook for each message. To send a message, the sender finds a codeword in the message subcodebook that is jointly typical with the given state sequence. The receiver decodes the codeword and hence finds the message. The details of the proof can be found, for example, in [2, Lecture 7].

* Chandra Nair is with the Chinese University of Hong Kong
† Abbas El Gamal and Yeow-Khiang Chia are with Stanford University

A straightforward extension of this Gelfand–Pinsker scheme to the DM-BC with DM state yields the lower bound on capacity

$$C \geq \max_{p(u|s)x(u,s)} \min\{I(U;Y_1) - I(U;S), I(U;Y_2) - I(U;S)\}. \tag{1}$$

In [3], it is conjectured that this rate is optimal in general. We show that this is not the case. We devise a new coding scheme for this channel that involves superposition coding, Marton coding, joint typicality encoding, and indirect decoding [4]. Our scheme yields the following lower bound on capacity.

Theorem 1: The common message capacity of the DM-BC with state information available non-causally at the sender is lower bounded by

$$\begin{aligned} C \geq \max\min\{&I(W,U;Y_1) - I(W,U;S),\ I(W,V;Y_2) - I(W,V;S), \\ &\frac{1}{2}\left(I(W,U;Y_1) - I(W,U;S) + I(W,V;Y_2) - I(W,V;S) - I(U;V|W,S)\right)\}, \end{aligned}$$

where the maximization is over distributions $p(w,u,v|s)$ and functions $x(w,u,v,s)$.

It is easy to see that this lower bound is at least as large as 1. We simply set $U = V = \emptyset$. We will show that our lower bound can in fact be strictly larger than 1.

In the following section, we formally define the problem of sending a common message over a DM-BC with DM state and describe the new coding scheme. In section III, we show through an example that the new lower bound can be strictly larger than the straightforward extension of the Gelfand-Pinsker result. In section IV, we present several classes of channels for which the new rate is optimum, including a class of compound Gaussian channels where the new rate achieves the dirty paper coding rate [5] for both channels simultaneously.

The notation used in this paper will follow that of El Gamal–Kim Lecture Notes on Network Information Theory [2, Lecture 1].

II. Achievability Scheme

Consider a 2-receiver DM-BC with DM state $(\mathcal{X}, \mathcal{S}, \{p(y_1, y_2|x, s)p(s), \mathcal{Y}_1, \mathcal{Y}_2)$ consisting of a finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, a finite state alphabet $\mathcal{S}$, two a collection of conditional pmfs $p(y_1, y_2|x, s)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$, and a pmf $p(s)$ on the state alphabet $\mathcal{S}$.

A $(2^{nR}, n)$ code for the DM-BC with noncausal state information available at the encoder consists of: (i) a message set $[1:2^{nR}]$, (ii) an encoder that assigns a codeword $x^n(m, s^n)$ to each message m and state sequence s^n, and (iii) two decoders, decoder 1 assigns an estimate $\hat{m}_1(y_1^n) \in [1:2^{nR}]$ or an error message e to each received sequence y_1^n and decoder 2 that assigns an estimate $\hat{m}_2(y_2^n) \in [1:2^{nR}]$ or an error message e to each received sequence y_2^n. We assume that M is uniformly distributed over $[1:2^{nR}]$. The probability of error is defined as $P_e^{(n)} = \mathrm{P}\{\hat{M}_1 \neq M \text{ or } \hat{M}_2 \neq M\}$.

A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity C is the supremum of all achievable rates.

The main result in this paper is the lower bound on the common message capacity of the DM-BC with DM state available non-causally at the encoder in Theorem 1. The proof of this theorem follows.

Codebook generation

- For each m, generate 2^{nT_0} $w^n(m, l_0)$ sequences according to $\prod_{i=1}^n p_W(w_i)$.
- For each (m, l_0) pair, generate 2^{nT_1} $u^n(m, l_0, l_1)$ sequences according to $\prod_{i=1}^n p_{U|W}(u_i|w_i)$.
- For each (m, l_0) pair, generate 2^{nT_2} $v^n(m, l_0, l_2)$ sequences according to $\prod_{i=1}^n p_{V|W}(v_i|w_i)$.

Encoding

The encoding procedure is illustrated in Figure 2.

- Given message m and state sequence s^n, the encoder finds $l_0 \in [1:2^{nT_0}]$ such that $(w^n(l_0), s^n) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one l_0, it chooses the smallest one. If there is none, it chooses $l_0 = 1$.

- The encoder next finds $l_1 \in [1:2^{nT_1}]$ and $l_2 \in [1:2^{nT_2}]$ such that $(w^n(m,l_0), s^n, u^n(m,l_0,l_1), v^n(m,l_0,l_2)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such pair, it chooses the pair with the smallest indices, first in l_1, then in l_2. If there is none, it chooses $(1,1)$.
- The encoder transmits $x(w_i, u_i, v_i, s_i)$ for $i \in [1:n]$.

Note that this scheme is essentially Marton coding with only diagonal product bins. Interestingly, the same encoding scheme can be used if we wish to send a common message M_0 to both receivers and private messages M_1 to Y_1 and M_2 to Y_2.

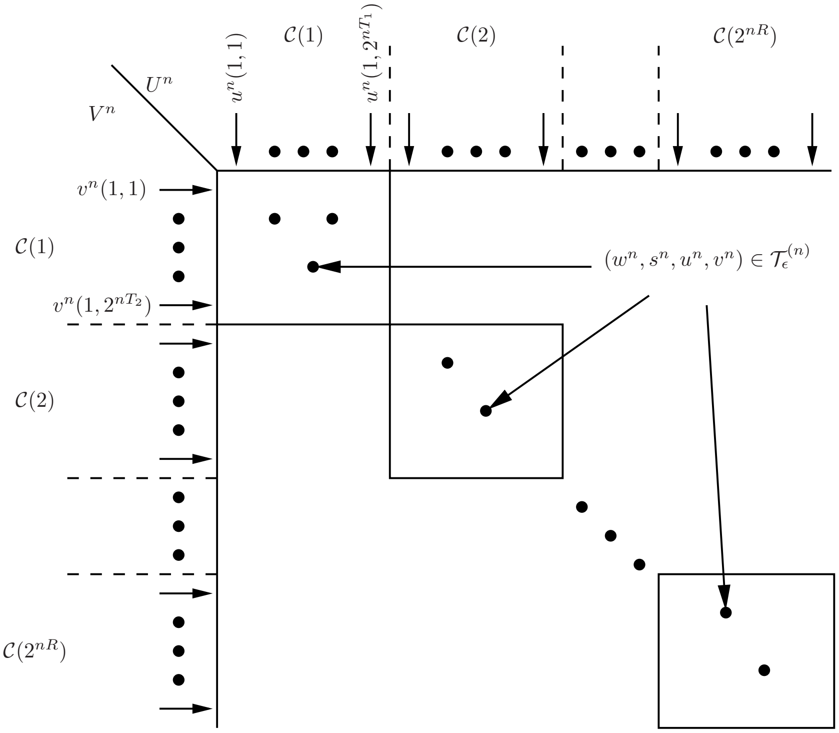

Fig. 2: Achievability scheme.

Decoding

Let $\epsilon' > \epsilon > 0$.

- Decoder 1 finds m indirectly by decoding (m, l_0). It declares that $\hat{m}_1$ is sent if it is the unique message such that $(w^n(\hat{m}_1, \hat{l}_0), u^n(\hat{m}_1, \hat{l}_0, \hat{l}_1), y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ for some $\hat{l}_0 \in [1:2^{nT_0}],\ \hat{l}_1 \in [1:2^{nT_1}]$.
- Decoder 2 finds m indirectly by decoding (m, l_0). It declares that $\hat{m}_2$ is sent if it is the unique message such that $(w^n(\hat{m}_2, \hat{l}_0), v^n(\hat{m}_2, \hat{l}_0, \hat{l}_2), y_2^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ for some $\hat{l}_0 \in [1:2^{nT_0}],\ \hat{l}_2 \in [1:2^{nT_2}]$.

Analysis of probability of error

An error may occur if either the encoder does not find a quadruple such that $(w^n(m,l_0), s^n, u^n(m,l_0,l_1), v^n(m,l_0,l_2)) \in \mathcal{T}_\epsilon^{(n)}$, or there is an error made by decoder 1 or 2.

We now analyze the probability of error averaged over codebooks. Without loss of generality, assume $M = 1$ is sent and (L_0, L_1, L_2) are the corresponding indices. Define the encoding error events

$$\mathcal{E}_{01} = \{(S^n, W^n(1,l_0)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } l_0\},$$
$$\mathcal{E}_{02} = \{(S^n, W^n(1,L_0), U^n(1,L_0,l_1), V^n(1,L_0,l_2)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } l_1, l_2\}$$

Then the total encoding error probability is

$$\mathrm{P}(\mathcal{E}_0) \leq \mathrm{P}(\mathcal{E}_{01}) + \mathrm{P}(\mathcal{E}_{02} \cap \mathcal{E}_{01}^c).$$

By the covering lemma [2, Lecture 3], the first term $\mathrm{P}(\mathcal{E}_{01}) \to 0$ as $n \to \infty$ if

$$T_0 > I(W;S).$$

Next, consider the second probability of error term

$$\begin{aligned}\mathrm{P}(\mathcal{E}_{02} \cap \mathcal{E}_{01}^c) &= \mathrm{P}\{(S^n, W^n(1,L_0), U^n(1,L_0,l_1), V^n(1,L_0,l_2)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } l_1, l_2\} \\ &\leq \sum_{(w^n,s^n)\in\mathcal{T}_\epsilon^{(n)}(W,S)} \mathrm{P}\{W^n(1,L_0) = w^n, S^n = s^n\}\, \mathrm{P}\{\mathcal{E}_{02}(s^n,w^n)\},\end{aligned}$$

where $\mathcal{E}_{02}(s^n,w^n)$ denotes the event that $\{(S^n = s^n, W^n(1,l_0) = w^n, U^n(1,L_0,l_1), V^n(1,L_0,l_2)) \notin \mathcal{T}_\epsilon^{(n)}\}$ for all l_1 and l_2, conditioned on the fact that the pair $(w^n,s^n) \in \mathcal{T}_\epsilon^{(n)}(W,S)$.

We show in Appendix A that $\mathrm{P}(\mathcal{E}_{02}(s^n,w^n)) \to 0$ as $n \to \infty$ if

$$\begin{aligned}T_1 &> I(U;S|W) + \delta(\epsilon), \\ T_2 &> I(V;S|W) + \delta(\epsilon), \\ T_1 + T_2 &> I(U;S|W) + I(V;S|W) + I(U;V|W,S) + \delta(\epsilon).\end{aligned}$$

Next consider the probability of decoding error. Consider the following error events for decoder 1

$$\begin{aligned}&\mathcal{E}_{11} = \{(S^n, W^n(1,L_0), U^n(1,L_0,L_1), Y_1^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\}, \\ &\mathcal{E}_{12} = \{(S^n, W^n(m,\tilde{l}_0), U^n(m,\tilde{l}_0,\tilde{l}_1), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)} \text{ for some } \tilde{l}_0 \in [1:2^{nT_0}],\ \tilde{l}_1 \in [1:2^{nT_1}],\ m \neq 1\}.\end{aligned}$$

The probability of error restricted to $\mathcal{E}_{01}^c$ for decoder 1 is upper bounded as

$$\mathrm{P}(\mathcal{E}_1) \leq \mathrm{P}(\mathcal{E}_{11} \cap \mathcal{E}_{01}^c) + \mathrm{P}(\mathcal{E}_{12}).$$

By the law of large numbers, the second term $\mathrm{P}(\mathcal{E}_{11} \cap \mathcal{E}_{01}^c) \to 0$ as $n \to \infty$. By the packing lemma [2, Lecture 3], the third term $\mathrm{P}(\mathcal{E}_{12}) \to 0$ as $n \to \infty$ if

$$R + T_0 + T_1 < I(W,U;Y_1) - \delta(\epsilon).$$

Similarly, the probability of error at decoder 2 tends to zero as $n \to \infty$ if

$$R + T_0 + T_2 < I(W,V;Y_2) - \delta(\epsilon).$$

Thus the overall probability of error tends to zero as $n \to \infty$ if

$$\begin{aligned}R + T_0 + T_1 &< I(W,U;Y_1), \\ R + T_0 + T_2 &< I(W,V;Y_2), \\ T_0 &> I(W;S), \\ T_1 &> I(U;S|W), \\ T_2 &> I(V;S|W), \\ T_1 + T_2 &> I(U;S|W) + I(V;S|W) + I(U;V|W,S).\end{aligned}$$

Performing Fourier-Motzkin Elimination on the stated rate constraints then gives the achievable rate stated in Theorem 1. □

Remarks:

1) It suffices to set X as a deterministic function of W and S in (1) and in Theorem 1. In (1), if X is a probabilistic mapping of (W,S), by the functional representation lemma [2] it can always be expressed as a function of (W,S,Q), where Q is independent of (W,S). Defining $W' = (W,Q)$, we obtain $X = x(W',S)$,

$I(W';Y_1) - I(W';S) \geq I(W;Y_1) - I(W;S)$ and $I(W';Y_2) - I(W';S) \geq I(W;Y_2) - I(W;S)$. Similar reasoning can also be applied to Theorem 1.

2) Theorem 1 can be readily extended to any finite number of receivers (equivalently, compound channel comprising a finite number of DMCs with DM state). In this case we have the common auxiliary random variable W and as many individual auxiliary random variables as the number of receivers.

III. Example

We now show through the example in Figure 3 that the achievable rate in Theorem 1 can be strictly larger than the rate achievable by the straightforward extension of the Gelfand-Pinsker coding scheme to the 2-receivers DM-BC with state given in 1, which we denote by R_{GP}.

We have $|\mathcal{X}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = |S| = 2$ and $\mathrm{P}\{S=0\} = 1/2$. The top half of the example corresponds to the channel transition probabilities when $S=0$ while the bottom half corresponds to the channel transition probabilities when $S=1$.

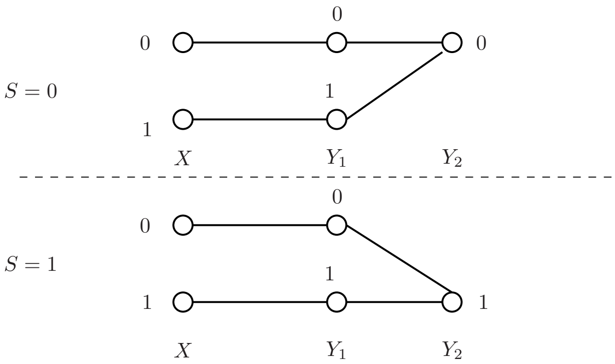

Fig. 3: Example DM-BC with DM state.

From Theorem 1, we set $W=\emptyset$, $U=Y_1$, $V=Y_2$ and $\mathrm{P}\{X=0|S=0\} = \mathrm{P}\{X=0|S=1\} = 0.5$. It is easy to verify that this choice of auxiliary random variables gives us an achievable rate of $R=0.5$. It is also clear that $C \leq I(X;Y|S) = H(Y|S) = 0.5$. Therefore, Theorem 1 achieves the common message capacity for this example.

A. $R_{\text{GP}} < C$

Expanding $I(U;Y_1) - I(U;S)$ in 1, we obtain

$$\begin{aligned} I(U;Y_1) - I(U;S) &= I(U;Y_1,S) - I(U;S) - I(U;S|Y_1) \\ &= H(Y_1|S) - H(Y_1|U,S) - I(U;S|Y_1) \\ &\leq H(Y_1|S) \leq \frac{1}{2}. \end{aligned}$$

To achieve $R_{\text{GP}} = H(Y_1|S)$, we require that $U \to Y_1 \to S$ form a Markov chain and Y_1 a function of (U,S). Since $Y_1 = X$ when $S=0$, we require that X is a function of U when $S=0$. Similarly, from $I(U;Y_2) - I(U;S)$, we require $U \to Y_2 \to S$ and Y_2 a function of (U,S). This implies that X is a function of (U,S). To further achieve $R_{\text{GP}} = 0.5$, we require that $\mathrm{P}\{X=0|S=0\} = \mathrm{P}\{X=0|S=1\} = 0.5$.

Let

$$\begin{aligned} \mathrm{P}\{U=i|X=0,S=0\} &= a_i, \\ \mathrm{P}\{U=i|X=1,S=0\} &= b_i, \\ \mathrm{P}\{U=i|X=0,S=1\} &= c_i, \\ \mathrm{P}\{U=i|X=1,S=1\} &= d_i. \end{aligned}$$

Since X is a function of (U, S), at least one of the two parameters a_i and b_i is equal to zero and at least one of c_i and d_i is also equal to zero. Further, from the Markov chain conditions $\mathrm{P}\{U = i|Y_2 = 0, S = 0\} = \mathrm{P}\{U = i|Y_2 = 0, S = 1\}$ and $\mathrm{P}\{U = i|Y_1 = 1, S = 0\} = \mathrm{P}\{U = i|Y_1 = 1, S = 1\}$, we obtain

$$\frac{a_i + b_i}{2} = c_i,$$
$$\frac{c_i + d_i}{2} = b_i.$$

If $a_i = 0$, $b_i = 2c_i$ and $d_i = 3c_i$. Since one of $c_i, d_i = 0$, this means that $a_i = b_i = c_i = d_i = 0$ or $\mathrm{P}\{U = i\} = 0$, which is a contradiction. Similarly, $b_i = 0$ forces $\mathrm{P}\{U = i\} = 0$, which is again a contradiction. This shows that there is no U with the required properties. Hence, $R_{\mathrm{GP}} < C$.

In fact, by means of a symmetrization argument given in Appendix B, we can show that R_{GP} can be computed exactly and is approximately equal to 0.41, implying a gap of 0.09 from C.

IV. Special Classes of Channels

Theorem 1 achieves the common message capacity in the following cases.

A. A class of deterministic channels with state

If both Y_1 and Y_2 are functions of (X, S) and $I(Y_1; Y_2|S) = 0$, then

$$C = \max_{p(x|s)} \min\{H(Y_1|S), H(Y_2|S)\}.$$

The example given in Section III belongs to this class of channels. Achievability follows from Theorem 1 by setting $W = \emptyset$, $U = Y_1$ and $V = Y_2$. The converse follows from the fact that $C \le \max_{p(x|s)} \min\{I(X; Y_1|S), I(X; Y_2|S)\}$.

Remark 1: One can also generalize this result to the class where Y_1 and Y_2 are functions of (X, S); Y_1 and Y_2 share common information (in the sense of Gäcs-Körner), i.e. there exists $Z = f(Y_1) = g(Y_2)$, and further $I(Y_1; Y_2|S, Z) = 0$. The achievability follows from Theorem 1 by setting $W = Z$, $U = Y_1$ and $V = Y_2$.

B. A class of compound Gaussian channels

We now develop a Gaussian analog of the example in Section IV. Let $S = (T, Z_S)$ where $T \sim \mathrm{Bern}(\alpha)$ and $Z_S \sim N(0, Q_T)$. The channel is defined as follows. When $T = 0$, we have

$$Y_1 = g_1 X + Z_S + Z_1,$$
$$Y_2 = 0,$$

where $Z_1 \sim N(0, 1)$. When $T = 1$, we have

$$Y_1 = 0,$$
$$Y_2 = g_2 X + Z_S + Z_2,$$

where $Z_2 \sim N(0, 1)$. The random variables $(T, Z_S), Z_1, Z_2$ are mutually independent. Since $Z_S \sim N(0, Q_T)$, we may have different variances in different states. Further, we assume an average transmit power constraint: $\sum_{i=1}^n \mathrm{E}(x_i^2(m, S^n)) \le nP,\ m \in [1 : 2^{nR}]$.

An upper bound on the capacity of this channel is

$$C \le \max_{p(x|s) \colon \mathrm{E}(X^2) \le P} \min\{I(X; Y_1|S), I(X; Y_2|S)\}.$$

It is easy to show that $I(X; Y_1|S) \le \alpha\, \mathrm{C}(g_1^2 P_1)$ and $I(X; Y_2|S) \le \bar{\alpha}\, \mathrm{C}(g_2^2 P_2)$, where $\alpha P_1 + \bar{\alpha} P_2 = P$ and $\mathrm{C}(P') = (1/2)\log(1 + P')$. From the writing on dirty paper result [5], in the single state case, the rate is $\mathrm{C}(P)$. Can we achieve the dirty paper coding rate for both Y_1 and Y_2 *simultaneously* for this more complicated class of compound Gaussian channels?

Using Theorem 1, we set $W = T$. When $T = 0$, we set

$$U = X_0 + \frac{g_1 P_1}{1 + g_1 P_1} Z_S, \text{ and } V = T,$$

where $X_1 \sim N(0, P_1)$. When $T = 1$, we set

$$U = T, \text{ and } V = X_1 + \frac{g_2 P_2}{1 + g_2 P_2} Z_S,$$

where $X_1 \sim N(0, P_2)$ and $\alpha P_1 + \bar{\alpha} P_2 = P$. This choice of random variables gives us the following achievable rate

$$\begin{aligned} R &< I(T;Y_1) + I(U;Y_1|T) - I(U;Z_S|T) - H(T), \\ R &< I(T;Y_2) + I(V;Y_2|T) - I(V;Z_S|T) - H(T), \\ 2R &< I(T;Y_1) + I(U;Y_1|T) - I(U;Z_S|T) - H(T) + \\ &\quad I(T;Y_2) + I(V;Y_2|T) - I(V;Z_S|T) - H(T) + I(U;V|T,Z_S). \end{aligned}$$

Since $I(T;Y_1) = I(T;Y_2) = H(T)$ and $I(U;V|T,Z_S) = 0$, simplifying the expression gives us

$$R < \max_{\alpha P_1 + \bar{\alpha} P_2 = P} \min\{\alpha C(g_1^2 P_1), \bar{\alpha} C(g_2^2 P_2)\},$$

which shows that we can achieve the dirty paper coding rate for both channels simultaneously.

V. Conclusion

We established a new achievable rate for the compound channel with DM state available noncausally at the encoder. The new achievable rate is shown to be strictly larger than the straightforward extension of the Gelfand-Pinsker coding scheme for a single state case. This result also implies that the straightforward extension of the Gelfand-Pinsker coding scheme for transmission over a DM-BC with DM state is not optimum.

References

[1] S. I. Gelfand and M. S. Psinker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[2] A. El Gamal and Y. H. Kim, "Lectures on network information theory," 2010, available online at ArXiv.

[3] P. Piantanida and S. Shamai, "Capacity of compound state-dependent channels with states known at the transmitter," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009, pp. 1968–1972.

[4] C. Nair and A. El Gamal, "The capacity region of a class of 3-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, October 2009.

[5] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

Appendix A
Bounding $\mathrm{P}(\mathcal{E}_{02}(s^n, w^n))$

The technique we use for bounding the term $\mathrm{P}(\mathcal{E}_{02}(s^n, w^n))$ is similar to that in the proof of the mutual covering lemma in [2, Lecture 9].

$\mathrm{P}(\mathcal{E}_{02}(s^n, w^n))$ is given by the probability of the event: $\{s^n, w^n, U^n(\tilde{l}_1), V^n(\tilde{l}_2)) \notin \mathcal{T}_\epsilon^{(n)}\}$ for all $\tilde{l}_1 \in [1:2^{nT_1}]$ and $\tilde{l}_2 \in [1:2^{nT_2}]$; where $U^n(\tilde{l}_1)$ and $V^n(\tilde{l}_2)$ are independently generated, conditioned on the given w^n, according to $\prod_{i=1}^n p_{U|W}(u_i|w_i)$ and $\prod_{i=1}^n p_{V|W}(v_i|w_i)$ respectively. Note that we are given $(s^n, w^n) \in \mathcal{T}_\epsilon^{(n)}$.

To show that $\mathrm{P}(\mathcal{E}_{02}) \to 0$ as $n \to \infty$, let $\mathcal{A} = \{(\tilde{l}_1, \tilde{l}_2) : (s^n, w^n, \tilde{U}^n(\tilde{l}_1), \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)}\}$ and $I(\tilde{l}_1, \tilde{l}_2) = 1$ if $(s^n, w^n, \tilde{U}^n(\tilde{l}_1), \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)}$ and 0 otherwise. Then, $|\mathcal{A}| = \sum_{\tilde{l}_1, \tilde{l}_2} I(\tilde{l}_1, \tilde{l}_2)$ and the expected number of jointly typical sequences is given by

$$\mathrm{E}\,|\mathcal{A}| = \sum_{\tilde{l}_1, \tilde{l}_2} \mathrm{P}\{(s^n, w^n, \tilde{U}^n(\tilde{l}_1), \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)}\}.$$

We further have the following bound on the probability:

$$\mathrm{P}\{(s^n, w^n, \tilde{U}^n(\tilde{l}_1), \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)}\}$$

$$\begin{aligned}
&= \sum_{\tilde{u}^n \in \mathcal{T}_\epsilon^{(n)}(\tilde{U}|w^n,s^n)} p(\tilde{u}^n)\,\mathrm{P}\{(s^n, w^n, \tilde{U}^n(\tilde{l}_1), \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)} | \tilde{U}^n(\tilde{l}_1) = \tilde{u}^n\} \\
&= \sum_{\tilde{u}^n \in \mathcal{T}_\epsilon^{(n)}(\tilde{U}|w^n,s^n)} \prod_{i=1}^n p_{U|W}(\tilde{u}_i|w_i)\,\mathrm{P}\{(s^n, w^n, \tilde{u}^n, \tilde{V}^n(\tilde{l}_2)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&\doteq \sum_{u^n \in \mathcal{T}_\epsilon^{(n)}(U|w^n,s^n)} 2^{-nH(U|W)} 2^{-nI(S,U;V|W)} \\
&\doteq 2^{-n(I(U;S|W)+I(S;V|W)+I(U;V|W,S))}.
\end{aligned}$$

Hence, we have

$$\mathrm{E}\,|\mathcal{A}| \geq 2^{n(T_1+T_2)} 2^{-n(I(U;S|W)+I(S;V|W)+I(U;V|W,S)+\delta(\epsilon))}.$$

Next, let

$$\begin{aligned}
p_1 &= \mathrm{P}\{(s^n, w^n, \tilde{U}^n(1), \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}\}, \\
p_2 &= \mathrm{P}\{(s^n, w^n, \tilde{U}^n(1), \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (s^n, w^n, \tilde{U}^n(1), \tilde{V}^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&= \sum_{\tilde{u}^n \in \mathcal{T}_\epsilon^{(n)}(U|w^n,s^n)} p(\tilde{u}^n)\,\mathrm{P}\{(s^n, w^n, \tilde{u}^n, \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}\}\,\mathrm{P}\{(s^n, w^n, \tilde{u}^n, \tilde{V}^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&\leq 2^{-n(I(U;S|W)+2I(V;S|W)+2I(U;V|W,S)-\delta(\epsilon))}, \\
p_3 &= \mathrm{P}\{(s^n, w^n, \tilde{U}^n(1), \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (s^n, w^n, \tilde{U}^n(2), \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&= \sum_{\tilde{v}^n \in \mathcal{T}_\epsilon^{(n)}(U|w^n,s^n)} p(\tilde{v}^n)\,\mathrm{P}\{(s^n, w^n, \tilde{v}^n, \tilde{U}^n(1)) \in \mathcal{T}_\epsilon^{(n)}\}\,\mathrm{P}\{(s^n, w^n, \tilde{v}^n, \tilde{U}^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&\leq 2^{-n(I(V;S|W)+2I(U;S|W)+2I(U;V|W,S)-\delta(\epsilon))}, \\
p_4 &= \mathrm{P}\{(s^n, w^n, \tilde{U}^n(1), \tilde{V}^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (s^n, w^n, \tilde{U}^n(2), \tilde{V}^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&= p_1^2.
\end{aligned}$$

Note that $\mathrm{E}\,|\mathcal{A}| = 2^{n(T_1+T_2)} p_1$.

$$\mathrm{E}\,|\mathcal{A}|^2 = 2^{n(T_1+T_2)} p_1 + \sum_{\tilde{l}_1,\tilde{l}_2} \sum_{\tilde{l}_2 \neq \tilde{l}_2'} p_2 + \sum_{\tilde{l}_1,\tilde{l}_2} \sum_{\tilde{l}_1 \neq \tilde{l}_1'} p_3 + \sum_{\tilde{l}_1,\tilde{l}_2} \sum_{\tilde{l}_1 \neq \tilde{l}_1'} \sum_{\tilde{l}_2 \neq \tilde{l}_2'} p_4.$$

Hence,

$$\mathrm{Var}(|\mathcal{A}|) \leq 2^{n(T_1+2T_2)} p_2 + 2^{n(2T_1+T_2)} p_3 + 2^{n(T_1+T_2)} p_1.$$

By Chebychev's inequality, we have

$$\begin{aligned}
\mathrm{P}\{|\mathcal{A}| = 0\} &\leq \mathrm{P}\{(|\mathcal{A}| - \mathrm{E}\,|\mathcal{A}|)^2 \geq (\mathrm{E}\,|\mathcal{A}|)^2\} \\
&\leq \frac{\mathrm{Var}(|\mathcal{A}|)}{(\mathrm{E}\,|\mathcal{A}|)^2} \\
&\leq 2^{-n(T_1 - I(U;S|W) - \delta(\epsilon))} + 2^{-n(T_2 - I(V;S|W) - \delta(\epsilon))} \\
&\quad + 2^{-n(T_1+T_2 - I(U;S|W) - I(V;S|W) - I(U;V|W,S) - \delta(\epsilon))}
\end{aligned}$$

Hence, $\mathrm{P}\{|\mathcal{A}| = 0\} \to 0$ as $n \to \infty$ if the following conditions are satisfied

$$\begin{aligned}
T_1 &> I(U;S|W) + \delta(\epsilon) \\
T_2 &> I(V;S|W) + \delta(\epsilon) \\
T_1 + T_2 &> I(U;S|W) + I(V;S|W) + I(U;V|W,S) + \delta(\epsilon).
\end{aligned}$$

Hence $\mathrm{P}(\mathcal{E}_{02}(s^n, w^n))$ goes to 0 as $n \to \infty$, provided the above conditions are satisfied.

APPENDIX B
EXACT EVALUATION OF R_{GP}

In this apendix, we evaluate R_{GP} using a symmetrization argument. Consider any (U,S,X) defined by $\mathrm{P}\{U=i,S=0\}=u_i, \mathrm{P}\{U=i,S=1\}=v_i, \mathrm{P}\{X=0|U=i,S=0\}=a_i, \mathrm{P}\{X=0|U=i,S=1\}=1-b_i$. From the fact that it suffices to look at $X=f(U,S)$, we have $a_i,b_i\in\{0,1\}$.

Then the following holds

$$\begin{aligned} H(Y) &= H\left(\textstyle\sum_i u_ia_i\right), & H(Y|U) &= \textstyle\sum_i(u_i+v_i)H\left(\frac{u_ia_i}{u_i+v_i}\right),\\ H(S) &= 1, & H(S|U) &= \textstyle\sum_i(u_i+v_i)H\left(\frac{u_i}{u_i+v_i}\right),\\ H(Z) &= H\left(\textstyle\sum_i v_ib_i\right), & H(Z|U) &= \textstyle\sum_i(u_i+v_i)H\left(\frac{v_ib_i}{u_i+v_i}\right). \end{aligned}$$

Now define a (U',S,X') (U' of size $2|\mathcal{U}|$) according to:

$$\begin{aligned} &\mathrm{P}\{U'=(i,1),S=0\}=u_i/2, \mathrm{P}\{U'=(i,2),S=0\}=v_i/2,\\ &\mathrm{P}\{X'=0|U'=(i,1),S=0\}=a_i, \mathrm{P}\{X'=0|U'=(i,2),S=0\}=b_i,\\ &\mathrm{P}\{U'=(i,1),S=1\}=v_i/2, \mathrm{P}\{U'=(i,2),S=1\}=u_i/2,\\ &\mathrm{P}\{X'=0|U'=(i,1),S=1\}=1-b_i, \mathrm{P}\{X'=0|U'=(i,2),S=1\}=b_i. \end{aligned}$$

Then observe that the new entropies are

$$\begin{aligned} H(Y') &= H\left(\sum_i \frac{u_ia_i}{2}+\frac{v_ib_i}{2}\right) \geq \frac{1}{2}(H(Y)+H(Z)),\\ H(Y'|U') &= \sum_i \frac{1}{2}(u_i+v_i)\left(H\left(\frac{u_ia_i}{u_i+v_i}\right)+H\left(\frac{v_ib_i}{u_i+v_i}\right)\right) = \frac{1}{2}(H(Y|U)+H(Z|U)),\\ H(S) &= 1,\\ H(S|U') &= \sum_i(u_i+v_i)H\left(\frac{u_i}{u_i+v_i}\right) = H(S|U),\\ H(Z') &= h(\sum_i \frac{u_ia_i}{2}+\frac{v_ib_i}{2}) \geq \frac{1}{2}(H(Y)+H(Z)),\\ H(Z'|U') &= \sum_i \frac{1}{2}(u_i+v_i)\left(H\left(\frac{u_ia_i}{u_i+v_i}\right)+H\left(\frac{v_ib_i}{u_i+v_i}\right)\right) = \frac{1}{2}(H(Y|U)+H(Z|U)). \end{aligned}$$

Thus, $I(U';Y')-I(U';S)=I(U';Z')-I(U';S)\geq \frac{1}{2}\left(I(U;Y)-I(U;S)+I(U;Z)-I(U;S)\right).$

A. Maximization of $I(U';Y')-I(U';S)$

Our maximization problem reduces to maximizing

$$I(U';Y')-I(U';S)$$

over all pmfs with the stated U' structure. That is, we wish to maximize

$$\begin{aligned} &H\left(\sum_i \frac{u_ia_i}{2}+\frac{v_ib_i}{2}\right) - \sum_i \frac{1}{2}(u_i+v_i)\left(H\left(\frac{u_ia_i}{u_i+v_i}\right)+H\left(\frac{v_ib_i}{u_i+v_i}\right)\right)\\ &\quad -1+\sum_i(u_i+v_i)H\left(\frac{u_i}{u_i+v_i}\right) \end{aligned}$$

subject to $\sum_i u_i = 0.5, \sum_i v_i = 0.5, a_i, b_i \in \{0,1\}$. The term can be rewritten as

$$H\left(\sum_i \frac{u_i a_i}{2} + \frac{v_i b_i}{2}\right) + \sum_i \frac{1}{2}(u_i+v_i)\left(H\left(\frac{u_i}{u_i+v_i}\right) - H\left(\frac{u_i a_i}{u_i+v_i}\right)\right)$$

$$-1 + \sum_i \frac{1}{2}(u_i+v_i)\left(H\left(\frac{v_i}{u_i+v_i}\right) - H\left(\frac{v_i b_i}{u_i+v_i}\right)\right).$$

Let $\mathcal{I}$ be the set of indices where $a_i = 0$ and $\mathcal{J}$ be the set of indices where $b_i = 0$. This implies that on $\mathcal{I}^c$ we have $a_i = 1$ and on $\mathcal{J}^c$ we have $b_i = 1$.

Thus, we wish to maximize

$$\begin{aligned} &H\left(\sum_{i\in\mathcal{I}^c} \frac{u_i}{2} + \sum_{i\in\mathcal{J}^c} \frac{v_i}{2}\right) + \sum_{i\in\mathcal{I}} \frac{1}{2}(u_i+v_i) H\left(\frac{u_i}{u_i+v_i}\right) \\ &-1 + \sum_{i\in\mathcal{J}} \frac{1}{2}(u_i+v_i) H\left(\frac{v_i}{u_i+v_i}\right). \end{aligned}$$

subject to $\sum_i u_i = 0.5, \sum_i v_i = 0.5$.

Define the following:

$$\begin{aligned} \frac{x_1}{2} &= \sum_{i\in\mathcal{I}\cap\mathcal{J}} u_i, \quad \frac{y_1}{2} = \sum_{i\in\mathcal{I}\cap\mathcal{J}} v_i, \\ \frac{x_2}{2} &= \sum_{i\in\mathcal{I}\cap\mathcal{J}^c} u_i, \quad \frac{y_2}{2} = \sum_{i\in\mathcal{I}\cap\mathcal{J}^c} v_i, \\ \frac{x_3}{2} &= \sum_{i\in\mathcal{I}^c\cap\mathcal{J}} u_i, \quad \frac{y_3}{2} = \sum_{i\in\mathcal{I}^c\cap\mathcal{J}} v_i, \\ \frac{x_4}{2} &= \sum_{i\in\mathcal{I}^c\cap\mathcal{J}^c} u_i, \quad \frac{y_4}{2} = \sum_{i\in\mathcal{I}^c\cap\mathcal{J}^c} v_i. \end{aligned}$$

Observe that $\sum_i x_i = 1, \sum_i y_i = 1$.

We note the following as a consequence of the concavity of the entropy function.

$$\begin{aligned} &\sum_{i\in\mathcal{I}} \frac{1}{2}(u_i+v_i) H\left(\frac{u_i}{u_i+v_i}\right) + \sum_{i\in\mathcal{J}} \frac{1}{2}(u_i+v_i) H\left(\frac{v_i}{u_i+v_i}\right) \\ &= \sum_{i\in\mathcal{I}\cap\mathcal{J}} (u_i+v_i) H\left(\frac{u_i}{u_i+v_i}\right) + \sum_{i\in\mathcal{I}\cap\mathcal{J}^c} \frac{1}{2}(u_i+v_i) H\left(\frac{u_i}{u_i+v_i}\right) + \sum_{i\in\mathcal{I}^c\cap\mathcal{J}} \frac{1}{2}(u_i+v_i) H\left(\frac{u_i}{u_i+v_i}\right) \\ &\le \frac{x_1+y_1}{2} H\left(\frac{x_1}{x_1+y_1}\right) + \frac{x_2+y_2}{4} H\left(\frac{x_2}{x_2+y_2}\right) + \frac{x_3+y_3}{4} H\left(\frac{x_3}{x_3+y_3}\right). \end{aligned}$$

Therefore we can upper bound the true maximum by the maximum of

$$H\left(\frac{x_3+x_4}{4} + \frac{y_2+y_4}{4}\right) + \frac{x_1+y_1}{2} H\left(\frac{x_1}{x_1+y_1}\right) + \frac{x_2+y_2}{4} H\left(\frac{x_2}{x_2+y_2}\right) + \frac{x_3+y_3}{4} H\left(\frac{x_3}{x_3+y_3}\right) - 1,$$

subject to $\sum_i x_i = 1, \sum_i y_i = 1$ and $x_i, y_i \ge 0$.

Now, we relax this maximization to $\sum_i x_i + y_i = 2$ and $x_i, y_i \ge 0$.

Define the partial sums $s_1 = x_1 + y_1$, $s_2 = x_2 + y_2$, $s_3 = x_3 + y_3$, and $s_4 = x_4 + y_4$. We re-write the maximization as

$$H\left(\frac{s_4}{4} + \frac{y_2+x_3}{4}\right) + \frac{s_1}{2} H\left(\frac{x_1}{s_1}\right) + \frac{s_2}{4} H\left(\frac{y_2}{s_2}\right) + \frac{s_3}{4} H\left(\frac{x_3}{s_3}\right) - 1,$$

subject to $0 \le x_1 \le s_1, 0 \le y_2 \le s_2, 0 \le x_3 \le s_3$ and $\sum_i s_i = 2$.

Using concavity of entropy, we can bound the maximum of the above expression by the maximum of

$$H\left(\frac{s_4}{4}+\frac{y_2+x_3}{4}\right)+\frac{s_1}{2}H\left(\frac{x_1}{s_1}\right)+\frac{s_2+s_3}{4}H\left(\frac{y_2+x_3}{s_2+s_3}\right)-1,$$

subject to $0 \leq x_1 \leq s_1, 0 \leq y_2+x_3 \leq s_2+s_3$ and $\sum_i s_i = 2$.

We first maximize with respect to x_1 and y_2+x_3 keeping the s_i terms fixed. Observe that the maximization is separable and it is concave in x_1 and y_2+x_3. Hence the maximum occurs when the first derivatives are zero; i.e. $x_1 = \frac{s_1}{2}$ and $\frac{s_4}{4}+\frac{y_2+x_3}{4} = 1-\frac{y_2+x_3}{s_2+s_3}$.

The second condition implies that

$$(y_2+x_3)(\frac{1}{4}+\frac{1}{s_2+s_3}) = 1-\frac{s_4}{4}, \text{ or } y_2+x_3 = \frac{(4-s_4)(s_2+s_3)}{4+s_2+s_3}.$$

Substituting for the optimal choices of x_1, y_2+x_3, the maximization reduces to that of

$$\frac{s_1}{2}+(1+\frac{s_2+s_3}{4})H(\frac{4-s_4}{4+s_2+s_3})-1,$$

subject to $\sum_i s_i = 2$, and $s_i \geq 0$.

Denote $s_2+s_3 = t$ and rewrite the maximization as

$$\left(1+\frac{t}{4}\right)H\left(\frac{4-s_4}{4+t}\right)-\frac{t}{2}-\frac{s_4}{2}$$

subject to $0 \leq t, 0 \leq s_4, s_4+t \leq 2$.

We divide into four cases:

1) The maximum is achieved at some strictly internal point, i.e. no inequality is tight.
2) The maximum is achieved when $t = 0$.
3) The maximum is achieved when $s_4 = 0$.
4) The maximum is achieved when $t+s_4 = 2$ but neither t or s_4 is zero.

It is not difficult to verify that the maximum over all four cases is attained by Case 3, with the setting $t = \frac{4}{3}, s_4 = 0$, and $s_1 = \frac{2}{3}$. The maximum value is approximately 0.41.

B. The maximizing $p(u,s), x(u,s)$

We now show all the relaxations can be made tight, i.e. there exists a suitable choice of U that achieves the derived bound.

Consider the following $|\mathcal{U}|$ with cardinality 3 defined according to:

$$\begin{aligned}
\mathrm{P}\{U=1,S=0\} &= \frac{1}{6}, \mathrm{P}\{X=0|U=1,S=0\}=0,\\
\mathrm{P}\{U=1,S=1\} &= \frac{1}{6}, \mathrm{P}\{X=1|U=1,S=1\}=0,\\
\mathrm{P}\{U=2,S=0\} &= \frac{1}{12}, \mathrm{P}\{X=0|U=2,S=0\}=0,\\
\mathrm{P}\{U=2,S=1\} &= \frac{1}{4}, \mathrm{P}\{X=1|U=2,S=1\}=1,\\
\mathrm{P}\{U=3,S=0\} &= \frac{1}{4}, \mathrm{P}\{X=0|U=3,S=0\}=1,\\
\mathrm{P}\{U=3,S=1\} &= \frac{1}{12}, \mathrm{P}\{X=1|U=3,S=1\}=0.
\end{aligned}$$

For this channel observe that

$$\begin{aligned}
I(U;Y)-I(U;S) &= I(U;Z)-I(U;S)\\
&= \frac{4}{3}H\left(\frac{3}{4}\right)-\frac{2}{3}\\
&\approx 0.41.
\end{aligned}$$