The Degraded Broadcast Channel with Non-Causal Action-Dependent Side Information

Yossef Steinberg
Dept. of EE, Technion, Israel
ysteinbe@ee.technion.ac.il

***Abstract*—In this paper we study the degraded broadcast channel (DBC) with non-causal side information that depends on encoder's actions. This is a continuation of previous works, by Steinberg and Weissman and by Ahmadi and Simeone, that dealt with the causal case. We first focus on the model where the actions depend only on the messages. Inner and outer bounds on the capacity region of the DBC with non-causal knowledge of action-dependent states are developed. For the special case where the stronger user is informed about the states, the bounds coincide and the capacity region is characterised.**

***Index Terms*—Action-dependent states; broadcast channel; non-causal side information.**

I. Introduction and Problem Definition

Let $\mathcal{A}$, $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be finite sets, and let $\mathcal{A}^n$ stand for the set of all n-sequences of letters from $\mathcal{A}$. An element $(a_1, a_2, \ldots, a_n)$ of $\mathcal{A}^n$ is denoted by a^n, with substrings written as $a_i^j = (a_i, \ldots, a_j)$. When the dimension n is clear from the context, boldface letters are used to denote n-vectors, thus $\boldsymbol{a}$ stands for a^n. Random variables are written by capital letters, with similar convention for random vectors: A^n, Y_i^j, $\boldsymbol{A}$ etc. For two jointly distributed random variables A and S, we denote by P_A, $P_{A,S}$, and $P_{S|A}$ the probability mass function (PMF) of A, their joint PMF, and the conditional PMF of S given A. A memoryless broadcast channel with action dependent states is a septuple $\{\mathcal{A}, \mathcal{S}, P_{S|A}, \mathcal{X}, P_{Y,Z|S,X}, \mathcal{Y}, \mathcal{Z}\}$ where $\mathcal{A}$ is the action alphabet, $\mathcal{S}$ the state space, $P_{S|A}$ a conditional PMF from the action alphabet to the state space, $\mathcal{X}$ the channel input alphabet, $\mathcal{Y}$ and $\mathcal{Z}$ the channel output alphabets, and $P_{Y,Z|S,X}$ a conditional PMF from $\mathcal{S} \times \mathcal{X}$ to $\mathcal{Y} \times \mathcal{Z}$. As a shorthand, and when the choice of the alphabets and $P_{S|A}$ is clear, we will refer to the broadcast channel just as $P_{Y,Z|S,X}$. Both the channel from action to state and that from state and channel input to channel outputs are assumed memoryless.

This work was supported by the ISRAEL SCIENCE FOUNDATION (grant no. 684/11). Part of this work was presented at the Information Theory and Applications Workshop, San Diego, CA, Feb. 9-15, 2013.

For notational convenience, the dimension in the PMF notation will sometimes be omitted. Thus $P_{S^n|A^n}(s^n|a^n)$ and $P_{Y^n,Z^n|S^n,X^n}(y^n, z^n|s^n, x^n)$ will be denoted by $P_{S|A}(\boldsymbol{s}|\boldsymbol{a})$ and $P_{Y,Z|S,X}(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{s}, \boldsymbol{x})$, respectively, and similarly for other tuples.

As in [8], we consider a vector cost function, of dimension $d \geq 1$, that takes into account input and action costs. Thus, let

$$\Lambda: \ \mathcal{A} \times \mathcal{X} \to [0, \infty)^d$$

be a single letter (vector-valued) cost function. The cost associated with vector of actions $\boldsymbol{a}$ and a channel input word $\boldsymbol{x}$ is

$$\Lambda(\boldsymbol{a}, \boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \Lambda(a_i, x_i).$$

Let $\Lambda_k(\boldsymbol{a}, \boldsymbol{x})$, $1 \leq k \leq d$, denote the kth coordinate of $\Lambda(\boldsymbol{a}, \boldsymbol{x})$. The cost function is called *separated*, and demoted by Λ^{sep}, if each of the components of Λ depends either only on the actions or only on the channel input, i.e.,

$$\begin{aligned} \Lambda_{k'}^{\text{sep}} &= \Lambda_{k'}^{\text{sep}}(\boldsymbol{a}), \quad 1 \leq k' \leq d', \\ \Lambda_{k}^{\text{sep}} &= \Lambda_{k}^{\text{sep}}(\boldsymbol{x}), \quad d'+1 \leq k \leq d, \end{aligned}$$

for some $0 \leq d' \leq d$. We denote the sets of messages by $\mathcal{N}_k = \{1, 2, \ldots, \nu_k\}$, $k = 1, 2$, where ν_1 and ν_2 are two integers.

A coding scheme for the broadcast channel with action dependent states available non-causally at the encoder operates as follows. A vector of actions $A^n = \alpha(M_1, M_2)$ is selected by the encoder based on the messages M_1 and M_2 which are drawn randomly from $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively. A random state vector S^n is now generated as the output of the memoryless channel $P_{S|A}$ with the action sequence A^n at its input. The encoder observes the state sequence in a non-causal manner and selects the channel input based on it and the pair of messages. In the description above the actions A^n depend only on the pair of messages M_1, M_2. A more general model, where the actions depend also on past values of the state, is termed *actions with feedback*, and was treated in [7] for the causal case. The model of actions with feedback for non-causal side information is not treated here.

The causal model was studied in [7] and independently in [1], [2]. We give next a formal definition of the coding scheme described above.

Definition 1. *An $(n, \nu_1, \nu_2, \lambda, \epsilon)$ code for the broadcast channel $P_{Y,Z|S,X}$ with action dependent states known causally at the encoder is an action mapping*

$$\alpha : \mathcal{N}_1 \times \mathcal{N}_2 \to \mathcal{A}^n, \tag{1}$$

an encoder mapping

$$f : \mathcal{S}^n \times \mathcal{N}_1 \times \mathcal{N}_2 \to \mathcal{X}, \tag{2}$$

and a pair of decoder maps

$$g_y : \mathcal{Y}^n \to \mathcal{N}_1$$
$$g_z : \mathcal{Z}^n \to \mathcal{N}_2$$

such that the average input costs are bounded by λ_k

$$\mathsf{E}[\Lambda_k(\boldsymbol{A}(M_1, M_2), f(\boldsymbol{S}, M_1, M_2))] \leq \lambda_k, \quad k = 1, 2, \ldots, d \tag{3}$$

and the probability of error P_e does not exceed ϵ. Here

$$P_e = 1 - \frac{1}{\nu_1 \nu_2} \sum_{m_1=1}^{\nu_1} \sum_{m_2=1}^{\nu_2} \sum_{\boldsymbol{s}} P_{S|A}(\boldsymbol{s}|\alpha(m_1, m_2)) P_{Y,Z|S,X}(g_y^{-1}(m_1), g_z^{-1}(m_2)|\boldsymbol{s}, f(\boldsymbol{s}, m_1, m_2)), \tag{4}$$

where $g_y^{-1}(m_1) \subset \mathcal{Y}^n$ and $g_z^{-1}(m_2) \subset \mathcal{Z}^n$ stand for the decoding sets of the messages m_1 and m_2, respectively. The rate pair (R_1, R_2) of the code is

$$R_1 = \frac{1}{n} \log \nu_1, \quad R_2 = \frac{1}{n} \log \nu_2.$$

A cost-rates triple (λ, R_1, R_2) is said to be achievable if for every $\epsilon > 0$ and sufficiently large n there exists an $(n, 2^{nR_1}, 2^{nR_2}, \lambda, \epsilon)$ code for the broadcast channel $P_{Y,Z|S,X}$ with non-causal action dependent states. The capacity-cost region for the channel is the closure of the set of all achievable cost-rates triples, and is denoted by $\mathcal{C}_{\text{nc}}$. The subscript nc stands for non-causal. For a given vector λ of input costs, $\mathcal{C}_{\text{nc}}(\lambda)$ stands for the section of $\mathcal{C}_{\text{nc}}$ at λ.

The probability of error (4) can be upper bounded by

$$P_e \leq P_{e,z} + P_{e,y} \tag{5}$$

where

$$P_{e,z} = 1 - \frac{1}{\nu_1 \nu_2} \sum_{m_1=1}^{\nu_1} \sum_{m_2=1}^{\nu_2} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{a}} P_{S|A}(\boldsymbol{s}|\alpha(m_1, m_2)) P_{Z|S,X}(g_z^{-1}(m_2)|\boldsymbol{s}, f(\boldsymbol{s}, m_1, m_2)) \tag{6}$$

and

$$P_{e,y} = 1 - \frac{1}{\nu_1 \nu_2} \sum_{m_1=1}^{\nu_1} \sum_{m_2=1}^{\nu_2} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{a}} P_{S|A}(\boldsymbol{s}|\alpha(m_1, m_2)) P_{Y|S,X}(g_y^{-1}(m_1)|\boldsymbol{s}, f(\boldsymbol{s}, m_1, m_2)), \tag{7}$$

re $P_{Y|S,X}$ and $P_{Z|S,X}$ are the conditional marginals derived from $P_{Y,Z|S,X}$. Therefore, the capacity region depends on $P_{Y,Z|S,X}$ only via its conditional marginals $P_{Y|S,X}$ and $P_{Z|S,X}$.

We use in the sequel the notion of degradedness as defined in [6]. A broadcast channel $P_{Y,Z|S,X}$ is said to be physically degraded if there exists a conditional PMF $P_{Z|Y}$ such that

$$P_{Y,Z|S,X}(y, z|s, x) = P_{Y|S,X}(y|s, x) P_{Z|Y}(z|y)$$

in which case Z is said to be the degraded component. A state dependent broadcast channel is said to be stochastically degraded if

$$P_{Z|S,X}(z|s, x) = \sum_y P_{Y|S,X}(y|s, x) P'_{Z|Y}(z|y)$$

for *some* conditional PMF $P'_{Z|Y}(z|y)$. As in the case where no actions are present [6], the capacity of the degraded broadcast channel with action-dependent states depends on the conditional PMF $P_{Y,Z|S,X}$ only via its marginals $P_{Y|S,X}$ and $P_{Z|S,X}$. Therefore, no distinction has to be made between physically degraded and stochastically degraded channels, and they will be termed degraded broadcast channels.

II. Main results

Let $\mathcal{P}$ be the set of all random variables $(A, A_2, K, U, S, X, Y, Z)$ such that A_2, K, and U take values in finite sets $\mathcal{A}_2, \mathcal{K}$, and $\mathcal{U}$, respectively, and

$$P_{A,A_2,K,U,S,X,Y,Z} = P_{A,A_2,K,U} P_{S|A} P_{X|A,A_2,K,U,S} \cdot P_{Y,Z|S,X}. \tag{8}$$

Denote by $\mathcal{R}_{\text{nc,i}}$ the set of all costs and rates (λ, R_1, R_2) such that

$$R_2 \leq I(K, A_2; Z) - I(K; A, S|A_2) \tag{9a}$$
$$R_1 \leq I(U; Y|K, A_2) - I(U; S|K, A_2, A) \tag{9b}$$
$$\mathsf{E}[\Lambda_k(A, X)] \leq \lambda_k, \quad k = 1, 2, \dots, d \tag{9c}$$

for some $P_{A,A_2,K,U,S,X,Y,Z} \in \mathcal{P}$. Note that the region $\mathcal{R}_{\text{nc,i}}$ remains unchanged if we replace (9b) with

$$R_1 \leq I(U, A; Y|K, A_2) - I(U; S|K, A_2, A) \tag{10}$$

since, on one hand, $I(U; Y|K, A_2) \leq I(U, A; Y|K, A_2)$, and on the other hand, we can always define a new external random variable $\tilde{U} = (U, A)$, which satisfies (8) and reduces (10) to (9b). The term $I(U; S|K, A_2, A)$ is not affected by this substitution due to the conditioning on A. The next proposition states a few properties of $\mathcal{R}_{\text{nc,i}}$.

Proposition 1. 1) *The region $\mathcal{R}_{\text{nc,i}}$ is convex*
2) *To exhaust $\mathcal{R}_{\text{nc,i}}$, it is enough to restrict $P_{A|A_2,K,U}$ and $P_{X|A,A_2,K,U,S}$ to be $0-1$ laws*
3) *To exhaust $\mathcal{R}_{\text{nc,i}}$, it is enough to restrict $\mathcal{A}_2$, $\mathcal{K}$, and $\mathcal{U}$ to satisfy*

$$\begin{aligned} |\mathcal{A}_2| &\leq |\mathcal{ASX}| + 1 \\ |\mathcal{K}| &\leq |\mathcal{ASX}| \left(|\mathcal{ASX}| + 1\right) + 1 \\ |\mathcal{U}| &\leq |\mathcal{ASX}| [|\mathcal{ASX}| \left(|\mathcal{ASX}| + 1\right) + 1] \\ &\quad \cdot [|\mathcal{ASX}| + 1] \end{aligned}$$

The proof is similar to the proof of [6, Proposition 4] and is omitted. We proceed to state the inner bound result.

Theorem 1. *For any discrete memoryless degraded broadcast channel with action-dependent states known non-causally at the encoder*

$$\mathcal{R}_{\text{nc,i}} \subseteq \mathcal{C}_{\text{nc}}.$$

A sketch of the proof is given in Section III. It is interesting to compare the region (9), and in particular (9a), with the capacity formula for single user channel with non-causal action-dependent states obtained by Weissman in [8]. The main difference is that here we have to perform superposition coding with respect to the actions, as reflected in (9a). Here the external random variable A_2 plays the role of the cenetr of the action cloud. Binning is performed not only with respect to the state S, but also with respect to the full action sequence A^n. This is because the actions A^n depend on the pair of messages (M_1, M_2), and thus in general cannot be decoded by the weaker user Z. Since it is not decoded by Z, binning must be performed with respect to the pair (A, S) in order to maintain joint typicality. We proceed to state the outer bound. Denote by $\mathcal{R}_{\text{nc,o}}$ the set of all rate pairs (R_1, R_2) and costs λ such that

$$R_2 \leq I(K, A_2; Z) - I(K; A, S|A_2) \tag{11a}$$
$$R_1 \leq I(U, A; Y|K) - I(U; S|K, A_2, A) \tag{11b}$$
$$R_1 + R_2 \leq I(U, K, A; Y) - I(U, K; S|A) \tag{11c}$$
$$\mathsf{E}[\Lambda_k(A, X)] \leq \lambda_k, \quad k = 1, \dots, d \tag{11d}$$

for some $P_{A,A_2,K,U,S,X,Y,Z} \in \mathcal{P}$.

Proposition 2. 1) *The region $\mathcal{R}_{\text{nc,o}}$ is convex*
2) *To exhaust $\mathcal{R}_{\text{nc,o}}$, it is enough to restrict $\mathcal{A}_2$, $\mathcal{K}$, $\mathcal{V}$, and $\mathcal{U}$ to satisfy*

$$\begin{aligned} |\mathcal{A}_2| &\leq |\mathcal{ASX}| + 1 \\ |\mathcal{K}| &\leq |\mathcal{ASX}| \left(|\mathcal{ASX}| + 1\right) + 2 \\ |\mathcal{U}| &\leq |\mathcal{ASX}| [|\mathcal{ASX}| \left(|\mathcal{ASX}| + 1\right) + 2] \\ &\quad \cdot [|\mathcal{ASX}| + 1] + 1 \end{aligned}$$

We next state the outer bound.

Theorem 2. *For any discrete memoryless degraded broadcast channel with action-dependent states known non-causally at the encoder*

$$\mathcal{C}_{\text{nc}} \subseteq \mathcal{R}_{\text{nc,o}}.$$

Due to space considerations, the proof is omitted. Theorems 1 and 2 above resemble parallel results on the broadcast channel with states - see [6, Theorems 1, 2]. The difference between the results here and in [6] is in the presence of actions, and the superposition coding and binning with respect to the actions, as discussed above.

The case where the stronger user Y is informed about the state, and in addition the cost function is separated, is a special case of the model treated here. It turns out that in this case the inner bound $\mathcal{R}_{\text{nc,i}}$ is tight, and admits a simpler form. Specifically, let $\mathcal{P}_{\text{nc}}$ be the collection of all (A, A_2, K, S, X, Y, Z) such that

$$P_{A,A_2,K,S,X,Y,Z} = P_{A,A_2,K} P_{S|A} P_{X|K,A_2,S} \cdot P_{Y,Z|S,X}. \tag{12}$$

and denote by $\mathcal{R}_{\text{nc}}$ the collection of all (R_1, R_2, λ) such that

$$R_2 \leq I(K, A_2; Z) - I(K; S|A_2) \tag{13a}$$
$$R_1 \leq I(A; S|A_2) + I(X; Y|S, K, A_2) \tag{13b}$$
$$\mathsf{E}\left[\Lambda_k^{\text{sep}}(A, X)\right] \leq \lambda_k, \quad k = 1, 2, \dots, d \tag{13c}$$

for some $P_{A,A_2,K,S,X,Y,Z} \in \mathcal{P}_{\rm nc}$. Similarly to the statement of Proposition 1, the following can be shown.

Proposition 3. 1) *The region $\mathcal{R}_{\rm nc}$ is convex*

2) *To exhaust $\mathcal{R}_{\rm nc}$, it is enough to restrict $\mathcal{A}_2$ and $\mathcal{K}$ to satisfy*

$$|\mathcal{A}_2| \le |\mathcal{A}\mathcal{S}\mathcal{X}| + 1$$
$$|\mathcal{K}| \le |\mathcal{A}\mathcal{S}\mathcal{X}| \left(|\mathcal{A}\mathcal{S}\mathcal{X}| + 1\right) + 1$$

The next theorem states our main result on the informed case.

Theorem 3. *For any discrete memoryless degraded broadcast channel with separated cost functions, and action-dependent states known non-causally at the encoder and the stronger decoder*

$$\mathcal{C}_{\rm nc} = \mathcal{R}_{\rm nc}.$$

III. Proofs

A. Proof of Theorem 1

A sketch of the proof of the achievability result of Theorem 1 is given here. Denote the messages of user Y (resp. Z) by m_1 (resp. m_2). We start by the random code construction. Pick a distribution from $\mathcal{P}$, and construct the corresponding random code as follows.

1) For every m_2, generate a sequence A_2^n i.i.d. according to P_{A_2}. Denote it by $A_2^n(m_2)$.
2) For every pair (m_1, m_2), generate A^n according to $\prod_{i=1}^n P_{A|A_2}(\cdot|A_{2,i}(m_2))$. Denote it by $A^n(m_1, m_2)$. This actions sequence generates the state S^n.
3) For every message m_2, generate a codebook $\{K^n(j, m_2)\}_{j=1}^{2^{nR_2'}}$ according to $\prod_{i=1}^n P_{K|A_2}(\cdot|A_{2,i}(m_2))$.
4) Let j_{m_2} the smallest integer j such that
$$(K^n(j, m_2), A^n(m_1, m_2), A_2^n(m_2), s^n) \in \mathcal{T}$$
Here $\mathcal{T}$ stands for the corresponding jointly typical set. The names of the random variables and the δ notation are dropped for brevity. By classical results, such an index j_{m_2} exists provided
$$R_2' > I(K; A, S|A_2) \tag{14}$$
5) For every pair (m_1, m_2), generate a codebook $\{U^n(l, m_1, m_2)\}_{j=1}^{2^{nR_1'}}$ according to $\prod_{i=1}^n P_{U|K,A_2,A}(\cdot|K_i^*, A_{2,i}(m_2), A_i(m_1, m_2))$ where $K^* = K^n(j_{m_2}, m_2)$.
6) Let l_{m_1,m_2} be the smallest integer l such that
$$(U^n(l, m_1, m_2), s^n, K^*, A_2^n(m_2), A^n(m_1, m_2)) \in \mathcal{T}$$
By classical results, such an index l_{m_1,m_2} exists provided
$$R_1' > I(U; S|K, A_2, A) \tag{15}$$
7) Finally, generate a codeword $X^n(m_1, m_2)$ randomly according to
$$\prod_{i=1}^n P_{X|U,K,A_2,A,S}(\cdot|U_i^*, K_i^*, A_{2,i}(m_2), A_i^*, s_i)$$
where $K^* = K^n(l_{m_1,m_2}, m_1, m_2)$ and $A^* = A^n(m_1, m_2)$.

Decoding. The degraded user Z looks for a pair (j, m_2) such that $(K^n(j, m_2), A_2^n(m_2), Z^n) \in \mathcal{T}$. By classical results, his decoding step succeeds with high probability if
$$R_2 + R_2' < I(K, A_2; Z) \tag{16}$$
Under this condition, the non-degraded user Y can also decode the pair (j, m_2). Denote by (j^*, m_2^*) his estimate of this pair. He now looks for a pair (l, m_1) satisfying
$$(U^n(l, m_1, m_2^*), A^n(m_1, m_2^*), K^n(j^*, m_2^*), A_2^n(m_2^*), Y^n) \in \mathcal{T} \tag{17}$$
This step succeeds with probability close to 1 when
$$R_1 + R_1' < I(U, A; Y|K, A_2) \tag{18}$$
The statement of the theorem follows from (14), (15), (16), and (18). □

B. Proof of Theorem 3

We start with the converse part. We are given a sequence of codes with increasing blocklength and vanishing probability of error. Denote by M_1 and M_2 the random messages intended to user Y and user Z, respectively. To prove the converse of Theorem 3, we can assume that the channel is physically degraded. Starting from the Fano inequality, we have

$$\begin{aligned}
&nR_2 - n\epsilon_n \le I(M_2; Z^n) \\
&= I(M_2; Z^n) - I(M_2; S^n|M_2) \\
&= \sum_{i=1}^n \left[I(M_2; Z_i|Z^{i-1}) - I(M_2; S_i|M_2, S_{i+1}^n) \right] \\
&= \sum_{i=1}^n \left[I(M_2 S_{i+1}^n; Z_i|Z^{i-1}) - I(S_{i+1}^n; Z_i|M_2 Z^{i-1}) \right)
\end{aligned}$$

$$
\begin{aligned}
&\quad - I(M_2Z^{i-1}; S_i|M_2, S_{i-1}^n) + I(Z^{i-1}; S_i|M_2S_{i+1}^n)\big] \\
&\overset{(a)}{=} \sum_{i=1}^n \big[I(M_2S_{i+1}^n; Z_i|Z^{i-1}) \\
&\quad -I(M_2Z^{i-1}; S_i|M_2, S_{i=1}^n)\big] \\
&\leq \sum_{i=1}^n \big[H(Z_i) - H(Z_i|Z^{i-1}M_2S_{i=1}^n) \\
&\quad -H(S_i|M_2, S_{i+1}^n) + H(S_i|M_2S_{i+1}^n Z^{i-1})\big] \\
&= \sum_{i=1}^n [I(K_i, A_{2,i}; Z_i) - I(K_i; S_i|A_{2,i})] \qquad (19)
\end{aligned}
$$

where $\lim_{n\to\infty} \epsilon_n = 0$, and (a) above is due to Csiszar's Lemma. In the last equality of (19) we used the definitions

$$A_{2,i} = M_2S_{i+1}^n \qquad (20)$$

$$K_i = M_2Z^{i-1} \qquad (21)$$

We bound next the rate for user 1. Here the state S is known to the decoder, hence starting with Fano inequality we obtain

$$
\begin{aligned}
&nR_1 - n\epsilon_n \leq I(M_1; Y^n|S^nM_2) \\
&= I(M_1; S^n|M_2) + I(M_1; Y^n|S^n, M_2) \\
&= \sum_{i=1}^n [I(M_1; S_i|M_2S_{i+1}^n) + I(M_1; Y_y|M_2S^nY^{i-1})] \\
&= \sum_{i=1}^n [I(M_1A_i; S_i|M_2S_{i+1}^n) + I(M_1; Y_i|M_2S^nY^{i-1})] \\
&\overset{(a)}{=} \sum_{i=1}^n [I(A_i; S_i|M_2S_{i+1}^n) \\
&\quad +I(M_1X_i; Y_i|M_2S^nY^{i-1}Z^{i-1}] \\
&\leq \sum_{i=1}^n [I(A_i; S_i|M_2S_{i+1}^n) \\
&\quad +I(M_1X_iS^{i-1}Y^{i-1}; Y_i|M_2S_i^nZ^{i-1})] \\
&\overset{(b)}{=} \sum_{i=1}^n [I(A_i; S_i|M_2S_{i+1}^n) + I(X_i; Y_i|M_2S_i^nZ^{i-1})] \\
&= \sum_{i=1}^n [I(A_i; S_i|A_{2,i}) + I(X_i; Y_i|S_iK_iA_{2,i})] \qquad (22)
\end{aligned}
$$

where (a) holds because of the Markov chain $S_i -\!\circ\!- A_i -\!\circ\!- (M_1, M_2, S_{i+1}^n)$ and because X_i is a function of the messages and the states, and (b) is true since given the input and the state at time i the channel output is independent of everything else. We also make use of the observation that due to the structure of (13) we can make X independent of A given K, A_2, S, without altering the region. The direct part follows from (19) and (22) by taking the limit of large n and the fact that the region stated in the theorem is convex.

Direct Part: We only describe briefly the differences between the proof of the direct part here and that of Theorem 1. There is no need to bin A_2 with respect to A, since A affects the channel only via S, which is known to the stronger user. Hence A signals directly via S. Due to lack of space, the details are omitted. □

References

[1] B. Ahmadi and O. Simeone, "On channels with action-dependent states," `http://arxiv.org/abs/1202.4438`, Feb. 2012.

[2] B. Ahmedi and O. Simeone, "On channels with action-dependent states," in *Proc. of IEEE Information Theory Workshop (ITW 2012)*, Lausanne, Switzerland, Sept. 3-7, 2012.

[3] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inf. Contr.*, vol. 46, pp. 1-15, 1980.

[4] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 379-381, May 1978.

[5] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp.289–293, 1958.

[6] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. IT-51, no. 8, pp. 2867–2877, Aug. 2005.

[7] Y. Steinberg and T. Weissman, "The degraded broadcast channel with action dependent states," in *Proc. of Int. Symposium on Information Theory (ISIT 2012*, Cambridge, MA, USA. July 1-6, 2012.

[8] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. IT-56, no. 11, pp. 5396–5411, No. 2010.